An Equilibrium-Dependent Retroviral mRNA Switch Regulates Translational Recoding

An equilibrium-dependent retroviral mRNA switch regulates translational recoding

Brian Houck-Loomis[1], Michael A. Durney[2], Carolina Salguero[2], Neelaabh Shankar[2], Julia M. Nagle[2], Stephen P. Goff [1] and Victoria M. D'Souza[2*]

[1]Department of Biochemistry and Molecular Biophysics, Howard Hughes Medical Institute, Columbia University, New York, NY. 10032. USA.

[2]Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA. 02138, USA.
[*]Correspondence: dsouza@mcb.harvard.edu

Most retroviruses require translational recoding of a viral mRNA stop codon to maintain a precise ratio of structural (Gag) and enzymatic (Pol) proteins during virus assembly[1,2]. Pol is expressed exclusively as a Gag-Pol fusion either by ribosomal frameshifting or by readthrough of the *gag* stop codon[3]. Both of these mechanisms occur infrequently and only affect 5-10% of translating ribosomes, allowing the virus to maintain the critical Gag to Gag-Pol ratio[4-8]. While it is understood that the frequency of the recoding event is regulated by *cis* RNA motifs, no mechanistic explanation is currently available for how the critical protein ratio is maintained. Here we present the NMR structure of the murine leukemia virus (MLV) recoding signal and show that a protonation-dependent switch occurs to induce the active conformation. The equilibrium is such that at physiological pH the active, readthrough permissive conformation is populated at ~6%: a level that correlates with *in vivo* protein quantities. The RNA functions by a highly sensitive, chemo-mechanical coupling tuned to ensure an optimal readthrough frequency. Similar observations for a frameshifting signal suggest that this novel equilibrium-based mechanism may have a general role in translational recoding.

Genetic recoding either by frameshifting or by readthrough requires *cis*-acting elements downstream of the recoding site in the mRNA. In MLV, a 63-nucleotide readthrough signal folds as a pseudoknotted[7-9] structure (MLV-PK) that can independently direct recoding of stop codons[5,6,10], even in heterologous RNA (see Supplementary Text and Supplementary Fig. 1a). Secondary structure mapping indicates that MLV-PK adopts a classic hairpin-type fold[9] consisting of two stems S1 and S2 connected by a single base loop L1 and an 18-nucleotide loop

L2 (Supplementary Fig. 1a). The eight base spacer between the UAG stop codon and the first base pair of stem S1 have been proposed to be unstructured[7,8,11]. While a majority of RNA pseudoknots contain specific tertiary interactions between stem and loop elements[12], our UV melting data indicated that these interactions in MLV-PK have an unexpected and pronounced dependence on pH (Supplementary Fig. 2). Assignment of the NMR data (Supplementary Figs. 3 and 4) unambiguously confirmed the presence of a pH-dependent equilibrium between two inter-converting conformations. Significant chemical shift perturbations are observed as the minor conformation becomes increasingly populated at lower pH values: this allowed us to select reporter residues, A17 and A14, by which to evaluate the equilibrium (Fig. 1b). As the N1 position of A17 becomes protonated, typical adjacent C2 chemical shift perturbations occur that report on the p*Ka* for this protonation event. Concurrent with protonation at A17, the C2 position of the A14 residue in the minor groove of stem S1 experiences a downfield chemical shift change (Fig. 1b) and we observe the appearance of nuclear Overhauser enhancements (NOEs) from the A14 H2 resonance to ribose resonances of G49 and U50 in loop L2. Together these data indicate that A14 enters a more electronegative environment upon approach of loop L2 and hence reports on formation of S1-L2 tertiary interactions. Interestingly, curve fitting for both the A17 and A14 resonances yielded a p*Ka* value of 6.23 and a p*fold* value of 6.20, respectively, implying that the two processes of protonation and tertiary structure formation may be coupled (Fig. 1c and Supplementary Fig. 5). We calculate that at physiological pH (7.4), a distribution would result in which ~6% of the MLV-PK population would have a tertiary structure while the remaining ~94% would lack the tertiary contacts. Since this distribution correlates both with our *in vivo* (Supplementary Fig. 6a) and with the previously observed *in vivo* levels of Gag-Pol and Gag[4-8], respectively, we asked if it is possible that an equilibrium ($PK_{active} \rightleftharpoons PK_{inactive}$) determines the Gag:Gag-Pol ratio.

To functionally test this hypothesis, we developed a dual luciferase reporter[13] *in vitro* translation system and modulated the response by varying the assay pH (Fig. 1d). A control experiment using a scrambled MLV-PK sequence showed that translational readthrough by ribosomes is not affected in the pH range tested. These data show that at pH 7.4 ~5.5% of ribosomal recoding occurs – a level in agreement with our measured value of conformational exchange from the NMR experiments. Strikingly, however, the activity of MLV-PK was found to be strictly

dependent upon the pH of the translation system, and exhibits enhanced readthrough activity at lower pH (Fig. 1d). Thus, these results indicate that formation of tertiary structure in MLV-PK leads to readthrough of the stop codon and furthermore that enriching the population with the tertiary fold enhances readthrough-stimulating activity. Mutational analysis to test the importance of protonation at A17 uncovered additional protonation sites, which demonstrate that MLV-PK is a complex multiple-proton sensor, and that protonation of A17 is one of two or more chemical triggers that govern the equilibrium (see Supplementary Text and Supplementary Figs. 6b and 7).

Although conformational heterogeneity precludes us from assigning all protonation sites and their effect on the global structure of PK_{active}, we can examine the role of $A17^{+}$ by obtaining a partial structure of PK_{active} (see Methods). Comparison with the complete $PK_{inactive}$ structure allows us to propose a mechanism for the conformational transitions that are required for readthrough activity (see Supplementary Figs. 8, 9, 10 and 11 and Supplementary Table 1). In both conformations a G-U wobble and three distinct A-U Watson-Crick base pairs are observed in the NMR data (Supplementary Fig. 8). While two canonical A-U base pairs are readily assigned to stem S1 (U12-A30 and A14-U27), the third pair can only be formed between bases U6 and A37. Thus, in contrast to previous chemical probing studies[9], our NMR assignments unambiguously prove that the helical stem S1 is extended by additional base pairing between residues G5-A8 in the spacer and loop L2 residues A34-U38 (Figs. 1a,e and Supplementary Fig. 1). The extended stem S1 helix also contains a 1×2 internal bulge[14,15] formed by bases A8, A34 and G35 (also see Supplementary Fig. 9). Additionally, our structures consistently show residue A29 in stem S1 to be looped out towards the major groove and not involved in tertiary interactions even in PK_{active} (Supplementary Fig. 10). A direct consequence of the 12 base pair (including the internal A8-G35 stack) extended stem S1 is a shortened four nucleotide spacer and a shortened loop L2 that has no long range NOEs to stem S1 and hence is flexible in $PK_{inactive}$ (Fig. 1e).

In the $PK_{inactive}$ structure (Fig. 1e, Supplementary Table 1 and Supplementary Fig. 10), the extended stem S1 coaxially stacks on stem S2 to form a collinear, quasi-continuous helix as evidenced by typical stacking NOEs between G16 and G52 across the helical junction

(Supplementary Fig. 4). Collinear stacking of stems S1 and S2 is possible because S2 contains seven base pairs which leaves an ~8 Å phosphorus-phosphorus distance for loop L1 across the S2 major groove. This distance can be spanned by the single nucleotide A17 without introducing any significant twist or bend at the helical junction[16]. However this coaxial stacking leads to steric hindrance at the junction, which, combined with shortened nature of loop L2, prevents S1-L2 interactions. In PK_{active}, protonation leads to changes that include a loss of stacking interactions between G16 and G52 at the helical junction, dramatic chemical shift change for residues at the S1-L2 turn (Supplementary Fig. 11) and the appearance of NOEs from G52 H2'/3', G52 H8 and G53 H8 to the protonated $A17^+$ H2 (Supplementary Fig. 12). These data suggest that in PK_{active} $A17^+$ forms an $A17^+$-C23:G53 base triple in the major groove (Fig. 1f). This repositioning of $A17^+$ decreases the length of loop L1 below the required minimal distance for coaxial stacking, and accordingly an inter-helical bend is introduced that relieves the steric hindrance at the junction and allows for S1-L2 tertiary interactions (Figs. 1f,g,h).

Since our data show that MLV-PK contains critical motifs that may regulate the structural transition (A17, A29, the helical junction and the S1-L2 turn) we hypothesized that manipulating these sites to favor tertiary contacts may modulate the equilibrium between PK_{active} and $PK_{inactive}$. Given that adenosine-stacking interactions are known to increase thermodynamic stability at turns[17], we asked if a U38A mutant would allow for increased A38 to A39 stacking at the S1-L2 turn and explain the previously observed increase in readthrough levels[8]. In addition, we used an A29C mutant to test if the looped out A29 sterically hinders S1-L2 triplex formation. Readthrough levels *in vivo* are significantly stimulated in both U38A (~80% increase) and A29C (~20%) (Supplementary Fig. 6c). Remarkably, the p*Ka* for A17 in U38A mutant is increased to 6.40 (Figs. 2a,b) and inter-helical bending occurs at higher pH (data not shown). Consequently this construct has an increased response to pH change (Fig. 2a). These data indicate that improved adenosine stacking in U38A facilities S1-L2 folding which is tightly coupled to the protonation event at the distal A17 site via the helical junction. Next, we asked if the two hyperactive mutants could have an additive effect in a U38A:A29C double mutant. Strikingly, a synergistic effect was observed for the double mutant affording a ~260% increase in readthrough levels *in vivo*, which correlates with a significantly elevated p*Ka* of 7.09 for A17 (Figs. 2b,c and Supplementary Fig. 13). Furthermore, we could also design mutants that disrupt the structural

integrity of the stem S1 and the S1-L2 turn and shift the equilibrium towards the $PK_{inactive}$ form: the pH response is thus attenuated (G15A:C26U) or abrogated (G11C, U6A, A39U) (Fig. 2a, Supplementary Figs. 6c and 11b).

Finally, we decided to engineer an artificial inter-helical bend by deleting one base pair in S2 (ΔGC) to test the importance of the mechanical response. Noticeably, our NMR data for ΔGC show that while the junction is unstacked at high pH, its ability to form the PK_{active} conformer is abrogated (Fig. 3a and Supplementary Fig. 14). This construct is not pH-responsive, indicating that the correct register for the S2-L1 triple helix formation does not form. We were able to rescue this mutant by using a ΔGC:U38A double mutant (Fig. 3b). Since U38A stabilizes the S1-L2 turn, this construct readjusts the inter-helical bend by bringing loop L2 in register with stem S1. This restores the pH-response and conclusively proves that long-range effects occur between the junction and the S1-L2 turn. While it was not possible to perform the *in vitro* translation experiment on the ΔGC mutant due the associated introduction of an in-frame $U_{38}AA$ stop codon, we could test the ΔGC:U38A construct (Fig. 3b). In agreement with our NMR data this construct rescues the pH-modulated recoding activity albeit with a lower efficiency compared to U38A (Fig. 3b). These experiments prove that the mechanical and chemical responses in MLV-PK are indeed coupled, and confirm our equilibrium-based RNA conformational switch model. Finally, similar pH-dependent, frameshift activity by the well-characterized frameshifting signal from Beet Western Yellow Virus[18,19] pseudoknot (Fig. 3d) suggests that fine-tuning a proton-driven structural equilibrium may be a common mechanism for regulating recoding of gene expression.

RNA is capable of a wide range of structural transitions ranging from subtle rearrangements to global folding that determine function[20]. MLV-PK is a highly responsive, multi-proton sensor that couples a chemical and mechanical response to induce a conformational transition. We present a novel equilibrium-based mechanism for regulation of the frequency of ribosomal recoding, wherein the only determinant required is an RNA conformational switch (Fig. 4). Previous explanations for how the recoding frequency is maintained have invoked mechanisms such as ribosomal mechanical stress, thermodynamic and kinetic stability of recoding signal, ribosomal pausing and ribosomal helicase unwinding[21,22]. These processes will play a role in

recoding but, according to our proposed mechanism, they would occur only if the translating ribosome encounters the active conformation of the RNA. Further support for our mechanism comes from the slow exchange lifetime of MLV-PK on the NMR time scale (>100 ms), which is slower than the known rate of ribosomal decoding *in vivo*[23]. This ensures that the pseudoknot tertiary fold will be maintained while a translating ribosome attempts to decode the Gag:Gag-Pol boundary.

METHODS SUMMARY

Detailed methods for RNA sample preparation, biophysical characterization[12,26], structure determination[26,27] and translational recoding assays[13] can be found in Methods.

REFERENCES

1 Felsenstein, K. M. & Goff, S. P. Expression of the gag-pol fusion protein of Moloney murine leukemia virus without gag protein does not induce virion formation or proteolytic processing. *J Virol* **62**, 2179-2182 (1988).

2 Shehu-Xhilaga, M., Crowe, S. M. & Mak, J. Maintenance of the Gag/Gag-Pol ratio is important for human immunodeficiency virus type 1 RNA dimerization and viral infectivity. *J Virol* **75**, 1834-1841 (2001).

3 Baranov, P. V., Gesteland, R. F. & Atkins, J. F. Recoding: translational bifurcations in gene expression. *Gene* **286**, 187-201 (2002).

4 Philipson, L. *et al.* Translation of MuLV and MSV RNAs in nuclease-treated reticulocyte extracts: enhancement of the gag-pol polypeptide with yeast suppressor tRNA. *Cell* **13**, 189-199 (1978).

5 Yoshinaka, Y., Katoh, I., Copeland, T. D. & Oroszlan, S. Murine leukemia virus protease is encoded by the gag-pol gene and is synthesized through suppression of an amber termination codon. *Proc Natl Acad Sci U S A* **82**, 1618-1622 (1985).

6 Wills, N. M., Gesteland, R. F. & Atkins, J. F. Evidence that a downstream pseudoknot is required for translational read-through of the Moloney murine leukemia virus gag stop codon. *Proc Natl Acad Sci USA* **88**, 6991-6995 (1991).

7 Wills, N. M., Gesteland, R. F. & Atkins, J. F. Pseudoknot-dependent read-through of retroviral gag termination codons: importance of sequences in the spacer and loop 2. *EMBO J* **13**, 4137-4144 (1994).

8 Alam, S. L., Wills, N. M., Ingram, J. A., Atkins, J. F. & Gesteland, R. F. Structural studies of the RNA pseudoknot required for readthrough of the gag-termination codon of murine leukemia virus. *J Mol Biol* **288**, 837-852(1999).

9 Brierley, I., Pennell, S. & Gilbert, R. J. Viral RNA pseudoknots: versatile motifs in gene expression and replication. *Nat Rev Microbiol* **5**, 598-610 (2007).

10 Honigman, A., Wolf, D., Yaish, S., Falk, H. & Panet, A. cis Acting RNA sequences control the gag-pol translation readthrough in murine leukemia virus. *Virology* **183**, 313-319 (1991).

11 Feng, Y. X., Yuan, H., Rein, A. & Levin, J. G. Bipartite signal for read-through suppression in murine leukemia virus mRNA: an eight-nucleotide purine-rich sequence

immediately downstream of the gag termination codon followed by an RNA pseudoknot. *J Virol* **66**, 5127-5132 (1992).

12 Nixon, P. L. & Giedroc, D. P. Energetics of a strongly pH dependent RNA tertiary structure in a frameshifting pseudoknot. *J Mol Biol* **296**, 659-671 (2000).

13 Grentzmann, G., Ingram, J. A., Kelly, P. J., Gesteland, R. F. & Atkins, J. F. A dual-luciferase reporter system for studying recoding signals. *RNA* **4**, 479-486 (1998).

14 Schroeder, S., Kim, J. & Turner, D. H. G.A and U.U mismatches can stabilize RNA internal loops of three nucleotides. *Biochemistry* **35**, 16105-16109 (1996).

15 Badhwar, J., Karri, S., Cass, C. K., Wunderlich, E. L. & Znosko, B. M. Thermodynamic characterization of RNA duplexes containing naturally occurring 1 x 2 nucleotide internal loops. *Biochemistry* **46**, 14715-14724 (2007).

16 Michiels, P. J. *et al.* Solution structure of the pseudoknot of SRV-1 RNA, involved in ribosomal frameshifting. *J Mol Biol* **310**, 1109-1123 (2001).

17 Clanton-Arrowood, K., McGurk, J. & Schroeder, S. J. 3' terminal nucleotides determine thermodynamic stabilities of mismatches at the ends of RNA helices. *Biochemistry* **47**, 13418-13427 (2008).

18 Cornish, P. V., Hennig, M. & Giedroc, D. P. A loop 2 cytidine-stem 1 minor groove interaction as a positive determinant for pseudoknot-stimulated -1 ribosomal frameshifting. *Proc Natl Acad Sci U S A* **102**, 12694-12699 (2005).

19 Atkins, J. F. & Gesteland, R. F. *Recoding : expansion of decoding rules enriches gene expression*. (Springer, 2010).

20 Al-Hashimi, H. M. & Walter, N. G. RNA dynamics: it is about time. *Curr Opin Struct Biol* **18**, 321-329 (2008).

21 Namy, O., Moran, S. J., Stuart, D. I., Gilbert, R. J. & Brierley, I. A mechanical explanation of RNA pseudoknot function in programmed ribosomal frameshifting. *Nature* **441**, 244-247 (2006).

22 Giedroc, D. P. & Cornish, P. V. Frameshifting RNA pseudoknots: structure and mechanism. *Virus Res* **139**, 193-208 (2009).

23 Agirrezabala, X. & Frank, J. Elongation in translation as a dynamic interaction among the ribosome, tRNA, and elongation factors EF-G and EF-Tu. *Q Rev Biophys* **42**, 159-200 (2009).

24 Kuhlman, B., Luisi, D. L., Young, P. & Raleigh, D. P. pKa values and the pH dependent stability of the N-terminal domain of L9 as probes of electrostatic interactions in the denatured state. Differentiation between local and nonlocal interactions. *Biochemistry* **38**, 4896-4903.

25 Legault, P. & Pardi, A. In-situ probing of adenine protonation in RNA by ^{13}C NMR. *J Am Chem Soc* **116**, 8390-8391 (1994).

26 D'Souza, V., Dey, A., Habib, D. & Summers, M. F. NMR structure of the 101-nucleotide core encapsidation signal of the Moloney murine leukemia virus. *J Mol Biol* **337**, 427-442 (2004).

27 Güntert, P., Mumenthaler, C. & Wüthrich, K. Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J Mol Biol* **273**, 283-298 (1997).

ACKNOWLEDGEMENTS

We thank J. Atkins for the p2luc plasmid, the New York Structural Biology Center for NMR time and Aaron Hawkins, Sara Leiman and John Gullinger for their contributions. B.H-L. and S.P.G would also like to thank the late Daniel Wolf for his inspiration and many helpful discussions. S.P.G acknowledges grant R37 CA30488 from the NCI/NIH and is an Investigator with the Howard Hughes Medical Institute.

AUTHOR CONTRIBUTIONS

V.M.D. and S.P.G. conceived of and designed the experiments. M.A.D., C.S., N.S., and B.H-L. did the structural analysis, B.H-L. did the recoding assays, and M.A.D and J.M.N. did the *in vivo* assay for the double mutant. M.A.D. and V.M.D. wrote the manuscript with assistance from B.H-L.

AUTHOR INFORMATION

Coordinates and restraints for the final ensemble of 10 structures of $PK_{inactive}$ have been deposited in the Protein Data Bank with ID code 2LC8. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to V.M.D. (dsouza@mcb.harvard.edu)

FIGURE LEGENDS:

Figure 1: MLV-PK conformational equilibrium-dependent readthrough.

a, NMR-derived secondary structure of MLV-PK. Spacer residues (bold grey) G5-A8 originally predicted to be unstructured are involved in stem S1 formation and consequently loop L2 is shortened. **b,** Adenosine C2 chemical shift changes in residues A17 and A14 during pH titrations (Supplementary Fig. 5). **c,** Non-linear data fits[24] for residues A17 and A14 to determine the p*Ka and* p*fold*, respectively (Supplementary Fig. 5). **d,** The effects of the translation pH on MLV-PK readthrough activity *in vitro*. Error bars indicate standard error (n=3). Decreasing the pH from 7.8 to 7.1 leads to a 500% increase in readthrough levels. At pH 7.0, our NMR data indicate that ~8% of MLV-PK is in the folded conformation, which correlates very closely to the *in vitro* readthrough level of ~7.5% at pH 7.1. **e,** NMR structure ensemble of $PK_{inactive}$ with residue A17 in orange and the flexible loop L2 in red. A 180º rotated view is shown in Supplementary Fig. 10a. **f,** Close-up views of the stem S2 and loop L1 major groove in both forms of MLV-PK and the $A17^{+}$-C23:G53 base-triple in PK_{active}. Dotted lines indicate hydrogen bond formation. **g,** Overlay of the $PK_{inactive}$ (red) and PK_{active} (green) structures showing the global transition upon PK_{active} formation. **h,** Schematic of the secondary structure model (bottom), a view from the helical axis (top). A17 protonation in MLV-PK overwinds the helical twist, which consequently leads to an interhelical bend and releases the steric hindrance at the S1-S2 junction.

Figure 2: Functional data for modulation of the equilibrium.

a, The effects of the translation pH on readthrough activity of MLV-PK S1-L2 turn mutants *in vitro*. The data show the positive effect of increased stacking in U38A, which is abrogated in U38C and correlates with *in vivo* activity (Supplementary Fig. 6c). Error bars indicate standard error (n=3). **b,** Hill plots[25] ($R \geq 0.99$) to determine the p*Ka* values of A17 in the indicated MLV-PK constructs. Δ_T is the total chemical shift difference for the C2 carbon of A17 in $PK_{inactive}$ and PK_{active} and Δ is the difference between the C2 chemical shift at a given pH and the maximum chemical shift. **c,** *In vivo* readthrough levels for the MLV-PK wild type (WT_B) and the U38A:A29C mutant.

Figure 3: Compensatory effects of S1-L2 turn on the engineered inter-helical bend and equilibrium-based frameshifting levels.

a, 1D NMR data at pH 7.4 (black) and pH 5.2 (blue) for the indicated MLV-PK constructs (left) to probe the long-range conformational transition. See Supplementary Fig. 2b for comparison with wild-type MLV-PK. Corresponding schematic topology models (right) indicate the degree of inter-helical bending and compensation by the S1-L2 turn. Broadening of the line widths are indicative of tertiary structure formation, which is abrogated in ΔGC but compensated in ΔGC:U38A mutant. **b,** The effects of the translation pH on the *in vitro* readthrough activity of the above MLV-PK constructs. Error bars indicate standard error (n=3). **c,** Comparison of the effects of the translation pH on MLV-PK readthrough and BWYV and HIV-1 frameshift activity *in vitro*. Error bars indicate standard error (n=3). In HIV-1, the recoding signal is understood to be a stem-loop[32] in which no protonation sites are predicted. In our pH-modulated assay the HIV-1 construct is not dependent on protonation for activity.

Figure 4: Model for equilibrium-based mechanism. Gene expression via translational recoding is regulated by a dynamic equilibrium ($PK_{active} \rightleftharpoons PK_{inactive}$) between an active, readthrough permissive conformation and an inactive, non-permissive conformation whose distribution determines the Gag:Gag-Pol ratio. Ribosomes that encounter the PK_{active} conformation continue translating through the stop codon. Structural and functional data strongly suggest that PK_{active} interacts with the ribosome before the stop codon is in the decoding center due to the shortened nature of the linker.

FULL METHODS:

RNA sample preparation. DNA sequences were designed to include the T7 promoter, an insert sequence corresponding to nucleotides 1-63 of MLV-PK, a *SmaI* linearization site and restriction sites for *BamHI* and *EcoRI* for insertion of the amplified product into pUC19. RNA samples for biophysical experiments were transcribed and purified as described[26]. Samples were prepared in various buffers as required for the experiments (see below).

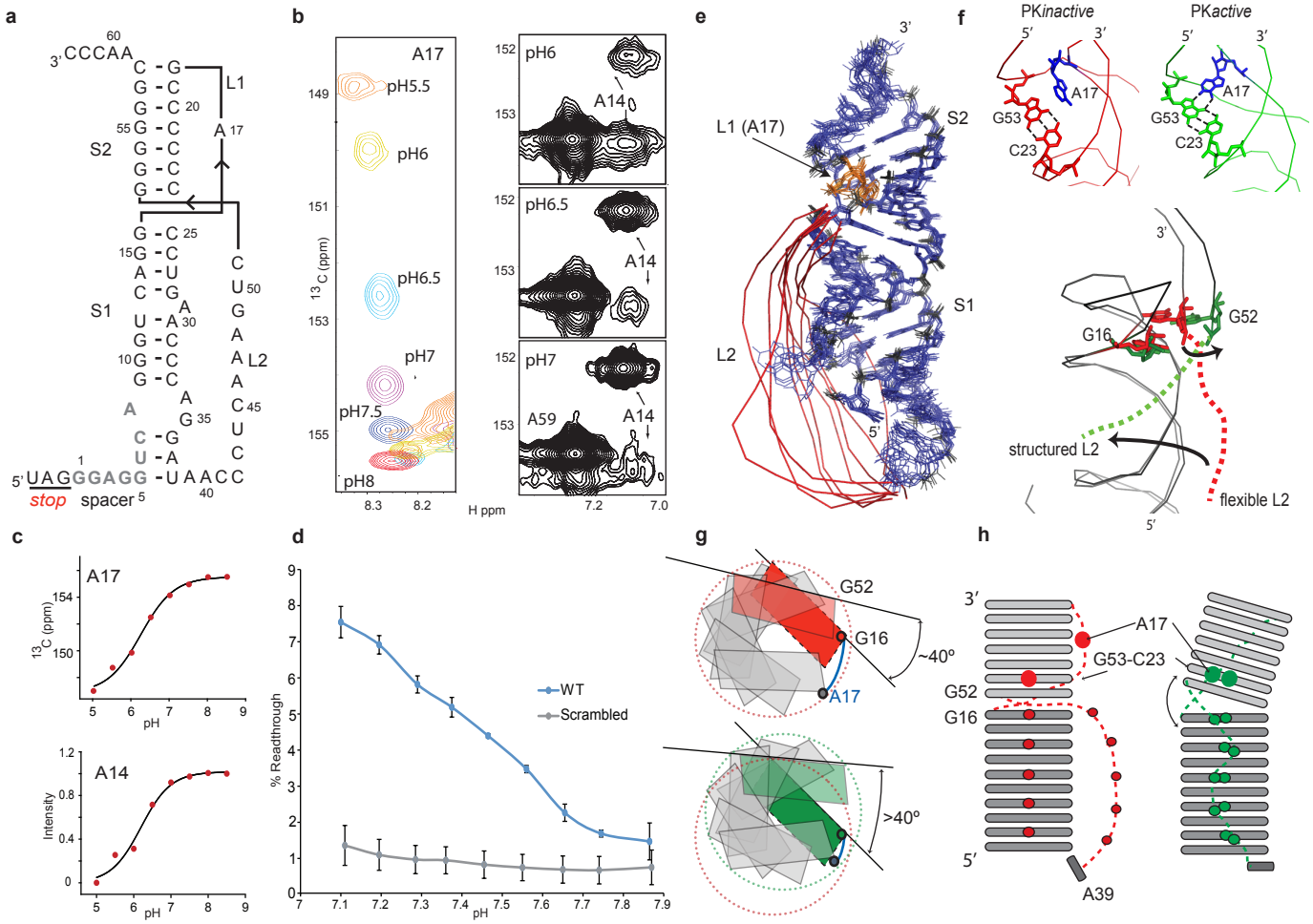
UV melting analysis. RNA samples for the wild-type, U38A and A39U MLV-PK constructs were annealed at 95°C under dilute conditions (1-2 μM) in UV buffer (20mM cacodylate, 1M NaCl). Absorbance versus temperature melting curves for the MLV-PK RNA samples were acquired at wavelengths of 260 nm and 280 nm. The heating rate was fixed at 0.5°C/minute with a Beckman DU800 single beam spectrophotometer equipped with a Peltier heating device. The pH experiments were performed at pH 5.5 and 7.5 and absorbance data were converted to first derivative versus temperature plots and analyzed as described previously[12].

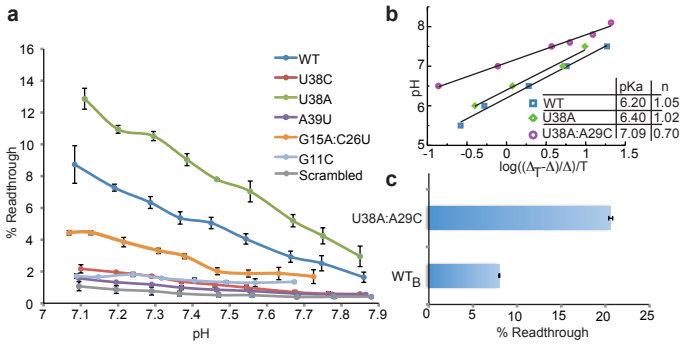
***In vivo* translation assay.** Readthrough was monitored using the p2luc dual luciferase reporter system[13]. The complete MLV-PK, BWYV and HIV-1 recoding sequence (MLV-PK: GACCCTAGATGACTAGGGAGGTCAGGGTCAGGAGCCCCCCCCTGAACCCAGGAAAC CCTCAAAGTCGGGGGGCAACCCGTC; BWYV: CAATTCATCGGGAAACTAAGTGCGC GGCACCGTCCGCGGAACAAACGGAAG; HIV-1:GAGACAGGCTAATTTTTTAGGGAAG ATCTGGCCTTCCCACAAGGGAAGGCCAGGAATTTTCTTCAGAGCAGACCATAGCC) were inserted between the renilla (Rluc) and firefly (Fluc) luciferase genes using either the *SalI* and *BamHI* or the *BamHI* and *SacI* restriction sites. To normalize for transfection and translation efficiency, control constructs were transfected in parallel to the sample plasmid. The control plasmid was identical to the MLV-PK construct except for a single U to C mutation in the stop codon and represented a readthrough efficiency of 100% at the mutated codon. The control constructs for BWYV and HIV-1 contained the same sequences in the -1 reading frame and the slippery sites were abrogated: (BWYV: CAATTCATC**C**GG**G**AA**G**CTAAGTGCGCGGCACCG TCCGCGGAACAAACGGAAG; HIV-1: GAGACAGGCTAA**C**TT**C**TT**A**AGGGAAGATCTG GCCTTCCCACAAGGGAAGGCCAGGGAATTTTCTTCAGAGCAGACCATAGCC). Dual

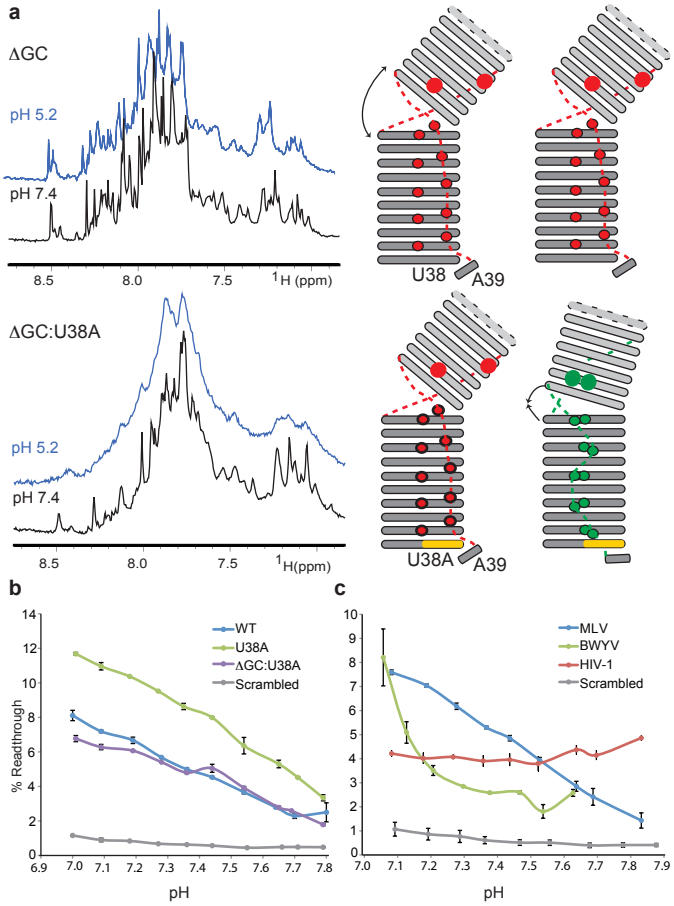
luciferase assays were performed in 96-well format. 293A cells (Invitrogen) were plated at a density of 2×10^4 cells/well in 125 μL DMEM plus penicillin and streptomycin and 10% FBS and grown overnight at 37°C in 5% CO_2. Transfections were performed with Fugene6 (Roche) and transfection mixes were made according to the manufacturer's instructions (50 ng reporter, 0.150 μL Fugene6, serum-free medium in a 17.5 μL final volume). Transfections were performed in triplicate and 5 μL of the transfection mix was added to each well. 24 hours post-transfection, the supernatant was decanted and 25 μL 1×PLB (passive lysis buffer, Promega) was added to each well. Lysates were incubated at room temperature with agitation for 10 minutes then subjected to one freeze-thaw cycle (−80°C). Dual luciferase measurements were performed with 20 μL of the lysate using a dual luciferase reporter assay system (Promega). The luciferase values from the control were used to normalize those from the sample, allowing us to report readthrough normalized to 100%.

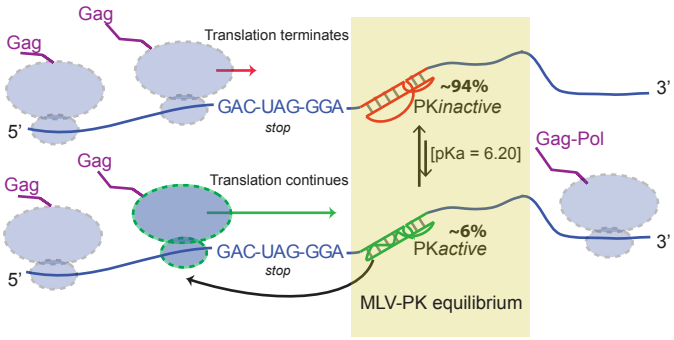
In vitro **translation assay.** Dual luciferase plasmids containing the MLV-PK recoding sequence inserted between the *SalI* and *BamHI* sites were linearized with *HpaI* and recovered by ethanol precipitation. The linearized template (1 μg) was used to synthesize capped RNA using the mMESSAGE mMACHINE kit (Ambion) and the manufacturer's protocol (2 hour incubation). The capped RNA was polyadenylated for 1 hour using the poly(A) tailing kit (Ambion) and the mRNA was purified using the MEGAClear kit (Ambion). RNA integrity was confirmed by running 1 μg on a denaturing glyoxal gel. The pH of the rabbit reticulocyte lysate (Ambion) was altered by the addition of dilute HCl and NaOH, the concentration of which was empirically determined. 5.75 μL of HCl (9.35, 6.32, 3.33, 0.0 mM) or NaOH (0.53, 1.30, 2.04, 2.78, 3.51, 4.23, 4.93, 5.63, 6.34 mM) was dispensed into 0.2 mL PCR tubes. Translation mixes (prepared in triplicate) were combined on ice and contained 1.25 μL 20× translation buffer (Ambion), 0.5 μL 50× (2.5 mM) methionine, 17 μL lysate and 0.5 μL (0.125 μg) reporter mRNA. The translation mix was added to the dilute acid or base, briefly vortexed, and incubated at 30°C for 1.5 hours. The reaction was quenched on ice for 10 minutes, 5 μL was removed for the dual luciferase assay and the pH of the remaining reaction was measured at 30°C with an InLab Micro pH probe (Mettler-Toledo).

NMR data acquisition, resonance assignment and structure calculations. For NMR experiments RNA samples were resuspended in NMR buffer (10mM Tris-HCl at pH 6.5 and pH 7.5, for PK_{active} and $PK_{inactive}$, respectively, and 10mM NaCl). NMR data were acquired using Bruker 700 MHz and 900 MHz spectrometers equipped with cryoprobes. Spectra were recorded at 298K and 308K with the exception of data for the imino region for which data were also recorded at 278K. Assignments for non-exchangeable ^{1}H and ^{13}C signals were obtained from 2D NOESY, 2D HMQC and 3D HMQC-NOESY datasets recorded with unlabeled and nucleotide-specific (AC^{CN}-MLV-PK, GC^{CN}-MLV-PK) selectively labeled samples and 2D NOESY samples obtained for nucleotide-specifically deuterated samples (GU^{H}-MLV-PK and AC^{H}-MLV-PK)[26]. Structures were calculated as described[26] using manually assigned restraints in CYANA[27]. The statistics table for the $PK_{inactive}$ structure ensemble is included in Supplementary Table 1. The changes that occur upon A17 protonation in stem S2, loop L1 and the helical junction were used to generate a model for this region of PK_{active}. The NOEs were confirmed with various MLV-PK mutants (U6A, A17U, A17C) (see Supplementary Figs 11 and 12). Molecular images were generated with PyMOL (www.pymol.org).


a
b
c
d
e
f
g
h
stop
spacer
L1
L2
S1
S2
A17
pH5.5
pH6
pH6.5
pH7
pH7.5
pH8
A14
A59
H ppm
pH
Intensity
% Readthrough
WT
Scrambled
L1 (A17)
PKinactive
PKactive
G53
C23
G16
G52
structured L2
flexible L2
~40°
>40°
G53-C23
A39


a
% Readthrough
pH
WT
U38C
U38A
A39U
G15A:C26U
G11C
Scrambled
b
pH
log((Δ_T−Δ)/Δ)/T
pKa
n
WT
6.20
1.05
U38A
6.40
1.02
U38A:A29C
7.09
0.70
c
U38A:A29C
WT_B
% Readthrough


a
ΔGC
pH 5.2
pH 7.4
8.5
8.0
7.5
1H (ppm)
U38
A39
ΔGC:U38A
pH 5.2
pH 7.4
8.5
8.0
7.5
1H(ppm)
U38A
A39
b
WT
U38A
ΔGC:U38A
Scrambled
% Readthrough
pH
c
MLV
BWYV
HIV-1
Scrambled
pH


Gag
Gag
Translation terminates
GAC-UAG-GGA
stop
5'
~94%
PKinactive
3'
[pKa = 6.20]
Gag-Pol
Gag
Gag
Translation continues
GAC-UAG-GGA
stop
5'
~6%
PKactive
3'
MLV-PK equilibrium

Supplementary Text

Introduction. All retroviruses contain three genes: namely, *gag*, *pol* and *env*, which code for structural, enzymatic and glycoprotein receptor proteins, respectively. In many retroviruses, translation cannot initiate autonomously in the *pol* gene, which is located immediately downstream of *gag*. Thus, whereas Gag can be independently translated, Pol is only produced as a Gag-Pol translational fusion[1]. This design has many advantages – it eliminates the need for additional promoter elements in an already compact genome and links the viral enzymes to the structural components, facilitating their incorporation into the virion during budding[3]. The key drawback is that the virus needs greater amounts of the structural proteins compared to enzymatic proteins for proper assembly of the viral particle[28]. In fact, retroviruses have to maintain a critical ratio of the Gag: Gag-Pol proteins for the formation of an infectious virus and manipulation of this ratio often results in non-infectious viral particles[1,2]. To circumvent this limitation retroviruses have evolved two mechanisms to maintain the required Gag:Gag-Pol ratio: (a) the translating ribosomes undergo a 'frameshift' so that the *gag* stop codon is not in frame and translation continues into the *pol* reading frame (e.g. human immunodeficiency virus, HIV) or (b) the translating ribosomes 'readthrough' the stop codon by inserting an amino acid instead of terminating translation (e.g. murine leukemia virus, MLV)[5,10] (see supplementary Fig. 1a). Both of these mechanisms occur infrequently and only affect about 5-10% of the translating ribosomes, allowing the virus to maintain the critical Gag to Gag-Pol ratio[4-8]. These expression strategies may have a broader utility in biology as frameshifting is known to be widely used in many organisms; in fact, the expression of ~148 genes in *D. melanogaster* are predicted to be regulated by readthrough events[29] (although only few have been biochemically characterized). There is currently no structure-based mechanism available for any readthrough-stimulating signal and consequently the basis of recoding frequency remains unclear.

MLV-PK contains multiple protonation sites. To test the importance of protonation at A17 we used NMR and the *in vitro* and *in vivo* recoding assays (Supplementary Fig. 7a). To this end, we constructed A17C, A17G and A17U mutants, expecting that only A17C would have a compensatory effect due to the ability of cytidine to become protonated at the N3 position. While A17C was indeed found to be fully functional and A17G has a significantly reduced activity,

A17U has an unexpected wild-type like activity in the *in vitro* assay (also see Supplementary Fig. 6b for *in vivo* activity). Guanosine and uridine residues cannot become protonated in the range tested yet both A17G and A17U are only capable of forming the tertiary interactions upon protonation. This finding was confirmed by NMR titrations, which show that tertiary structure formation in these mutants is still sensitive to pH. Although these mutants directly report on additional protonation events in MLV-PK, line broadening in the NMR spectra preclude us from determining the exact number of events. Nevertheless, the A17G data proves the importance of protonation in loop L1, and in addition show that in this construct the equilibrium is perturbed. More importantly however, since the *in vitro* readthrough levels for A17U are identical throughout the range tested compared to the wild-type our data provide strong evidence that the additional protonation events occur simultaneously with A17 protonation, and that all sites may have similar p*Ka* values. Accordingly, the A17 data in the wild-type fit well to the simplest model of a single, non-interacting pH titration site with a Hill coefficient of 1.05 (Fig. 2b) indicating that it titrates independently from other residues.

The contrasting results of A17U and A17G are intriguing and suggest that the long-range effects of loop L1 protonation are compensated or decoupled from tertiary structure formation in A17U. The RNA structural switch model can be used to understand these unexpected results. An A17C substitution allows for a similar C^{+}17-C23:G53 triple base as the wild-type and hence accounts for the wild-type like activity in the *in vitro* translation assay. NMR analysis of A17U shows that this residue is outside the major groove due to the increased electrostatic repulsion and leads to unstacking of the helical junction (Supplementary Fig. 7b). This pre-unstacked junction in A17U allows for the tertiary S1-L2 interaction to occur upon protonation of the other site(s), thus decoupling loop L1 from S1-L2 and allowing for wild type like activity. In A17G, we hypothesize that the electrostatics of the guanosine base would allow for a major groove placement similar to adenine but since no protonation occurs, it cannot trigger the unstacking of the stems S1 and S2. Most importantly this inefficient coupling of stems S1 and S2 perturbs the equilibrium and inhibits PK_{active} formation. Both of these mutants confirm that coaxial stacking at the helical junction is exceptionally sensitive to the precise placement of residue A17. These data also conclusively prove that MLV-PK is a complex multiple-proton sensor, and that protonation of A17 is only one of the chemical triggers that communicates a mechanical

rearrangement in global conformation via the helical junction.

MLV-PK mutant U38A:A29C has increased readthrough activity. In this mutant the S1-L2turn is predominantly in the folded conformation (Supplementary Fig. 13). Interestingly, our NMR data show that substituting the looped out A29 unstacks the junction and removes the steric hindrance allowing for the loop L2 A48-C51 segment to interact with stem S1. Thus, this region of MLV-PK is in the PK_{active} conformation even at high pH and accordingly the pKa of A17 is significantly elevated to 7.09 (Figs. 2b). However, residues at the S1-L2 turn in this construct are in slow exchange on the NMR time scale and thus populate the two distinct PK_{active} and $PK_{inactive}$ conformations, requiring protonation to shift the equilibrium towards PK_{active}. Strikingly, a synergistic effect was observed for the double mutant affording a ~260% increase in readthrough levels *in vivo*, (Fig. 2c). In other words, the double mutant is capable of inducing readthrough in one out of every five translation events.

References:

28 Karacostas, V., Wolffe, E. J., Nagashima, K., Gonda, M. A. & Moss, B. Overexpression of the HIV-1 gag-pol polyprotein results in intracellular activation of HIV-1 protease and inhibition of assembly and budding of virus-like particles. *Virology* **193**, 661-671 (1993).

29 Stark, A. *et al.* Discovery of functional elements in 12 *Drosophila* genomes using evolutionary signatures. *Nature* **450**, 219-232 (2007).

30 Case, D. A. *et al.* The Amber biomolecular simulation programs. *J Comput Chem* **26**, (2005).

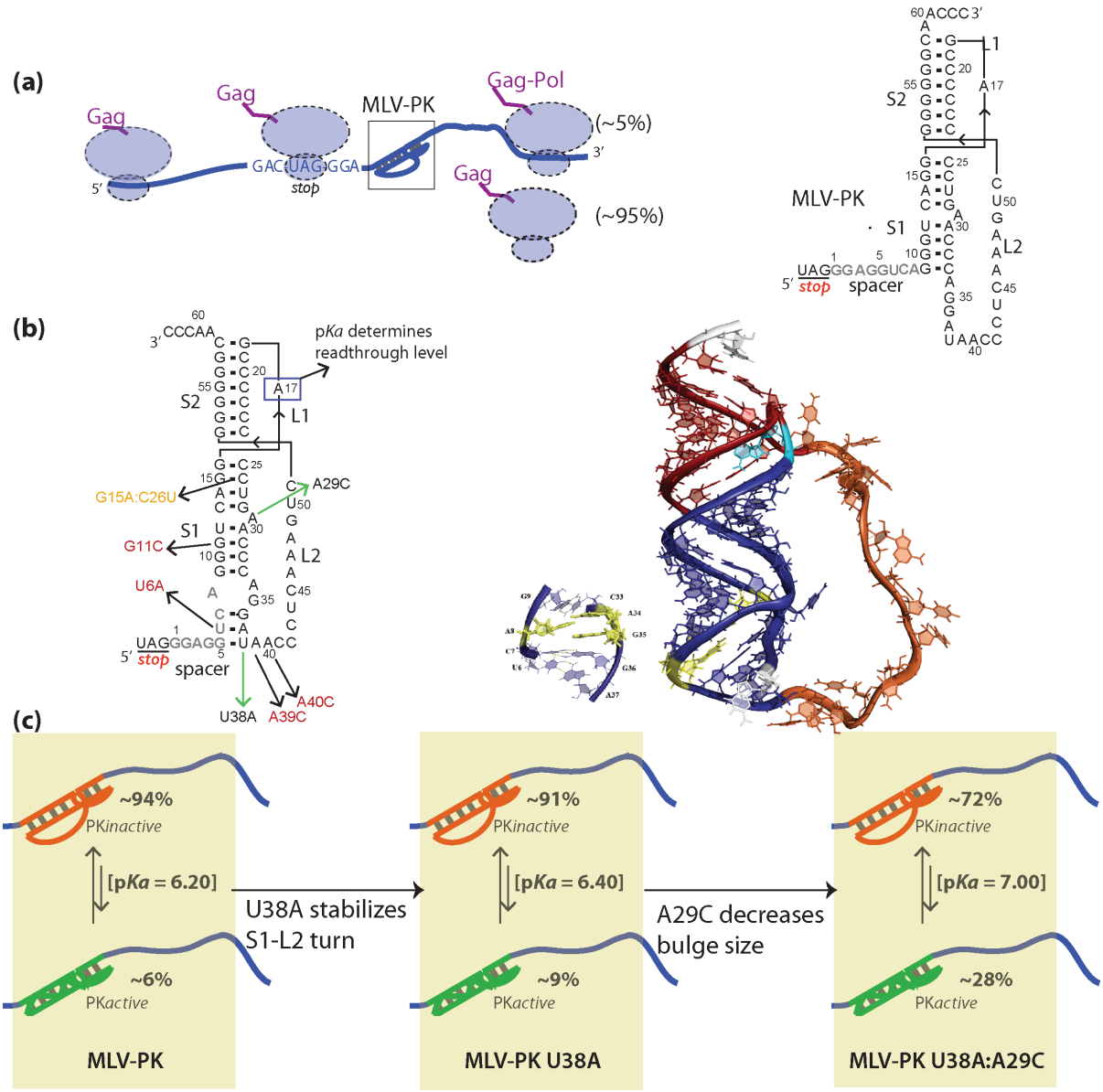
31 Hansen, M. R., Mueller, L. & Pardi, A. Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nat Struct Biol* **5**, 1065-1074, (1998).

32 Staple, D. W. & Butcher, S. E. Solution structure and thermodynamic investigation of the HIV-1 frameshift inducing element. *J Mol Biol* **349**, 1011-1023 (2005).

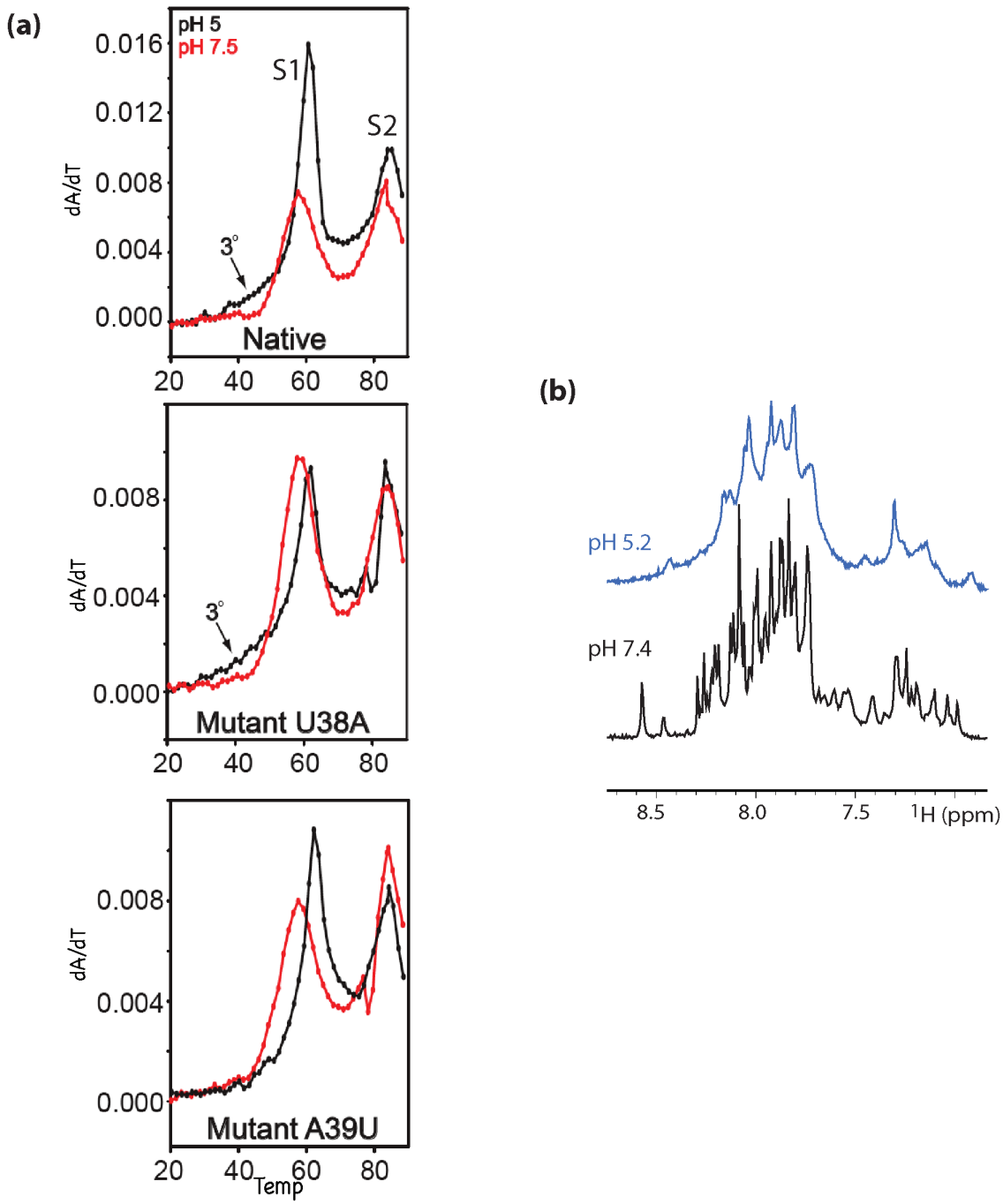
Table 1: NMR and refinement statistics for the final PKinactive structures

NMR distance and dihedral restraints	
Total upper and lower limit distance restraints	673
Total NOE	390
Intra-residue	267
Inter-residue	123
Sequential (|i-j| = 1)	113
Non-sequential (|i-j| > 1)	10
Hydrogen bonds	200
Based on A-form geometry	83
Total dihedral angle restraints	284
Structure statistics	
Target function (mean ± s.d., $Å^2$)	**0.14 ± 0.02**
Violations (mean ± s.d.)	
Distance restraints (> 2Å)	0.00 ± 0.00
Dihedral angle restraints (>5°)	0.00 ± 0.00
Max. dihedral angle violation (°)	0.45 ± 0.42
Max. upper distance restraint violation (Å)	0.08 ± 0.05
Max. lower distance restraint violation (Å)	0.06 ± 0.00
Average pairwise r.m.s.d.** (Å)	
Base paired heavy atoms and A17	**0.59 ± 0.13**

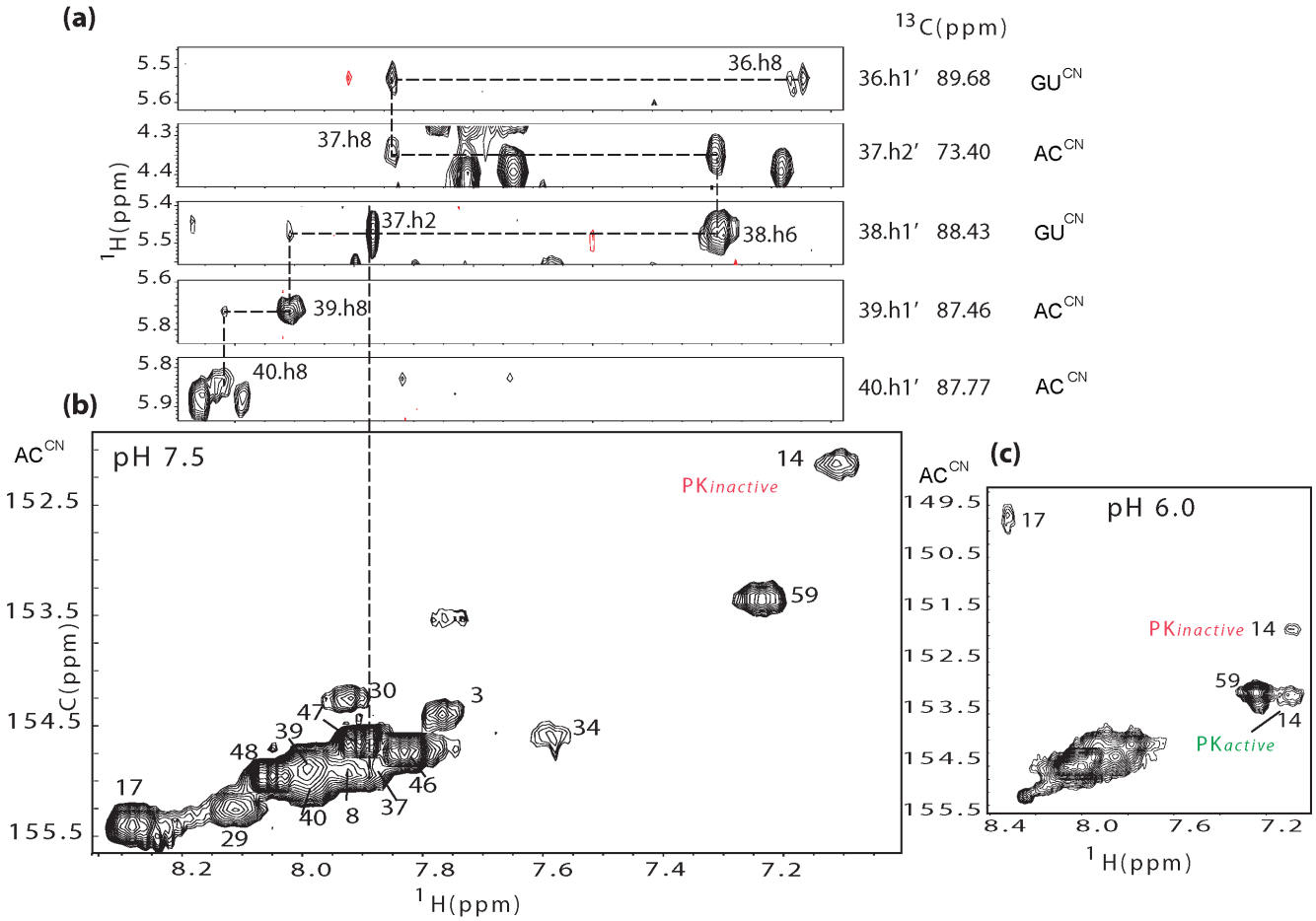
**Pairwise r.m.s.d. to the mean structure was calculated among 10 refined structures, which represent 5% of the starting structures.



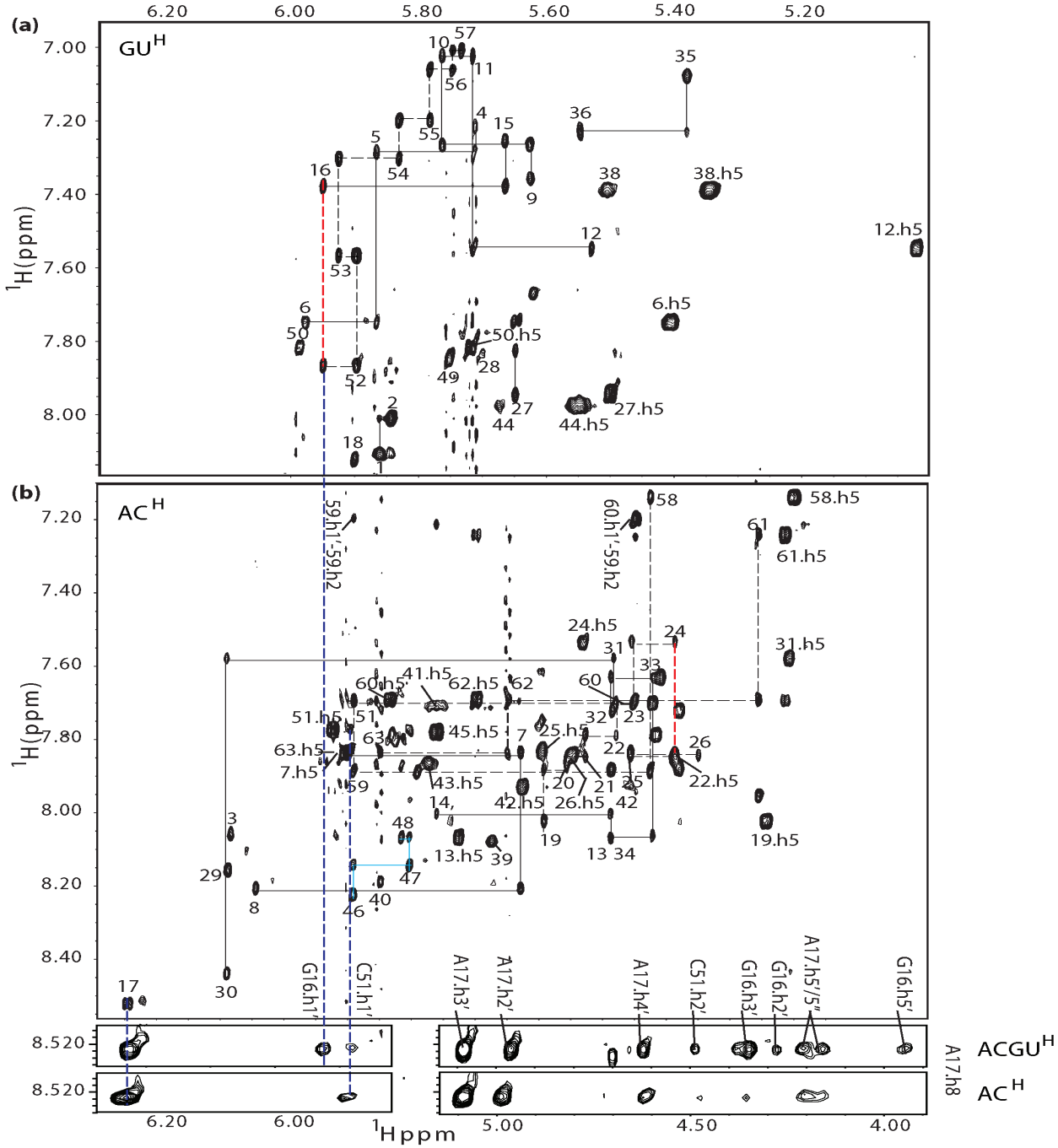
Supplementary Figure 1: Overview of the main findings. a, Schematic representation of ribosomal readthrough (***Note:*** Also see Supplementary Text) and secondary structure model of MLV-PK based on chemical modification and mutational analysis and prior to this study. **b,** Secondary and tertiary structure of MLV-PK based on current NMR studies and summary of MLV pseudoknot mutations. Mutations in green stimulate an increase in readthrough activity; mutations in orange display a mild to moderate decrease in readthrough; and mutations in red effectively abolish the readthrough-stimulating activity of the pseudoknot. **c,** Using NMR and functional assays, it was found that the MLV-PK is able to adopt a readthrough-permissive conformation, which involves tertiary structure formation, in a protonation-dependent manner. Most importantly, the p*Ka* of 6.2 for the transition between the two conformers indicates that only ~6% of MLV pseudoknots are folded into a readthrough-permissive structure at physiological pH. It exists predominantly in the inactive, non-permissive conformation by maintaining sub-optimal turns, steric hindrances and protonation requirements to afford an ~6% stop codon readthrough rate. To prove the equilibrium model, structure guided mutants that perturb the conformational equilibrium and enrich for the readthrough permissive conformer were made. For example, the U38A mutation increases readthrough efficiency 80%, and A29C:U38A double mutant by 260%, respectively (U38A stabilizes the S1-L2 turn and A29C decreases steric hindrance). These levels correlate with the population distributions seen in our NMR samples and the p*Ka* of A17.



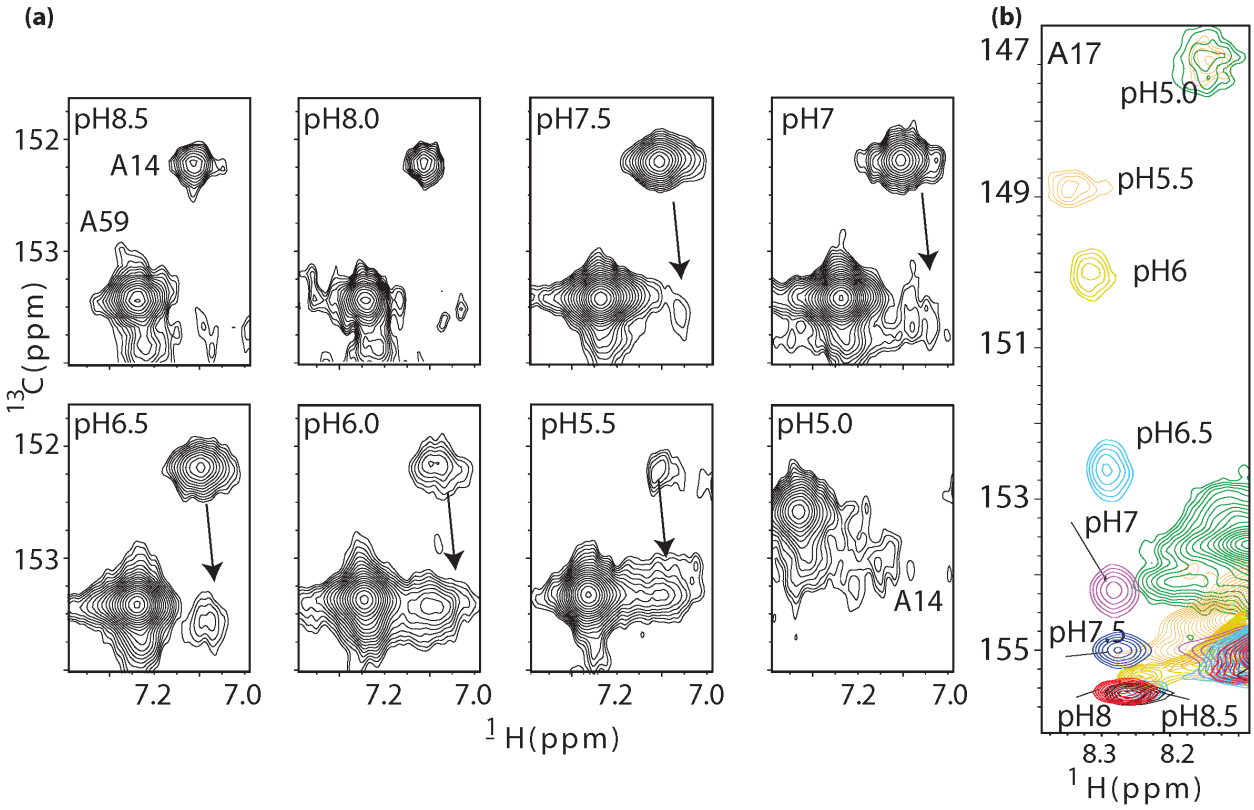
Supplementary Figure 2: Biophysical data showing tertiary structure formation in MLV-PK. a, A majority of RNA pseudoknot structures show the presence of triplex formation via extensive and specific hydrogen bonding resulting in correct register formation between stem and loop elements. Unfolding of pseudoknots is therefore characterized by typical UV transitions that comprise of an initial melting of the tertiary structure and a subsequent melting of secondary structure[12]. UV melting experiments for MLV-PK constructs at pH 7.5 (red) and pH 5.5 (black) uncover unexpected pH dependence for formation of tertiary structure. The tertiary structure transitions are indicated by arrows and melting events for stems S1 and S2 are labeled. MLV-PK mutant A39U does not form the tertiary structure and is hence inactive (see Fig. 2a and Supplementary Fig. 6c) **b,** 1D NMR data at pH 7.4 (black) and pH 5.2 (blue) for MLV-PK showing broadening of line widths that are diagnostic of tertiary structure formation.



Supplementary Figure 3: Assignment of MLV-PK using heteronuclear NMR data. Regions of the 3D HMQC-NOESY spectrum **a,** obtained for the specifically labeled AC^{CN}-MLV-PK and GU^{CN}-MLV-PK samples and 2D HMQC **b,** data for the adenine residues in a AC^{CN}-MLV-PK labeled sample. The assignment strategy (see Methods) allowed for unambiguous assignment of all residues at pH 7.5. A standard A-helical walk is obtained for residues 36-38 that confirms an extended Stem S2. This was further verified by the imino proton data (see Supplementary Fig. 8) **c,** 2D HMQC data for the AC^{CN}-MLV-PK labeled sample at pH 6.0. Although chemical shift changes are observed for most of the adenine C2-H2 groups due to tertiary structure formation, characteristic carbon chemical shift perturbation for base protonation was observed only for residue A17.

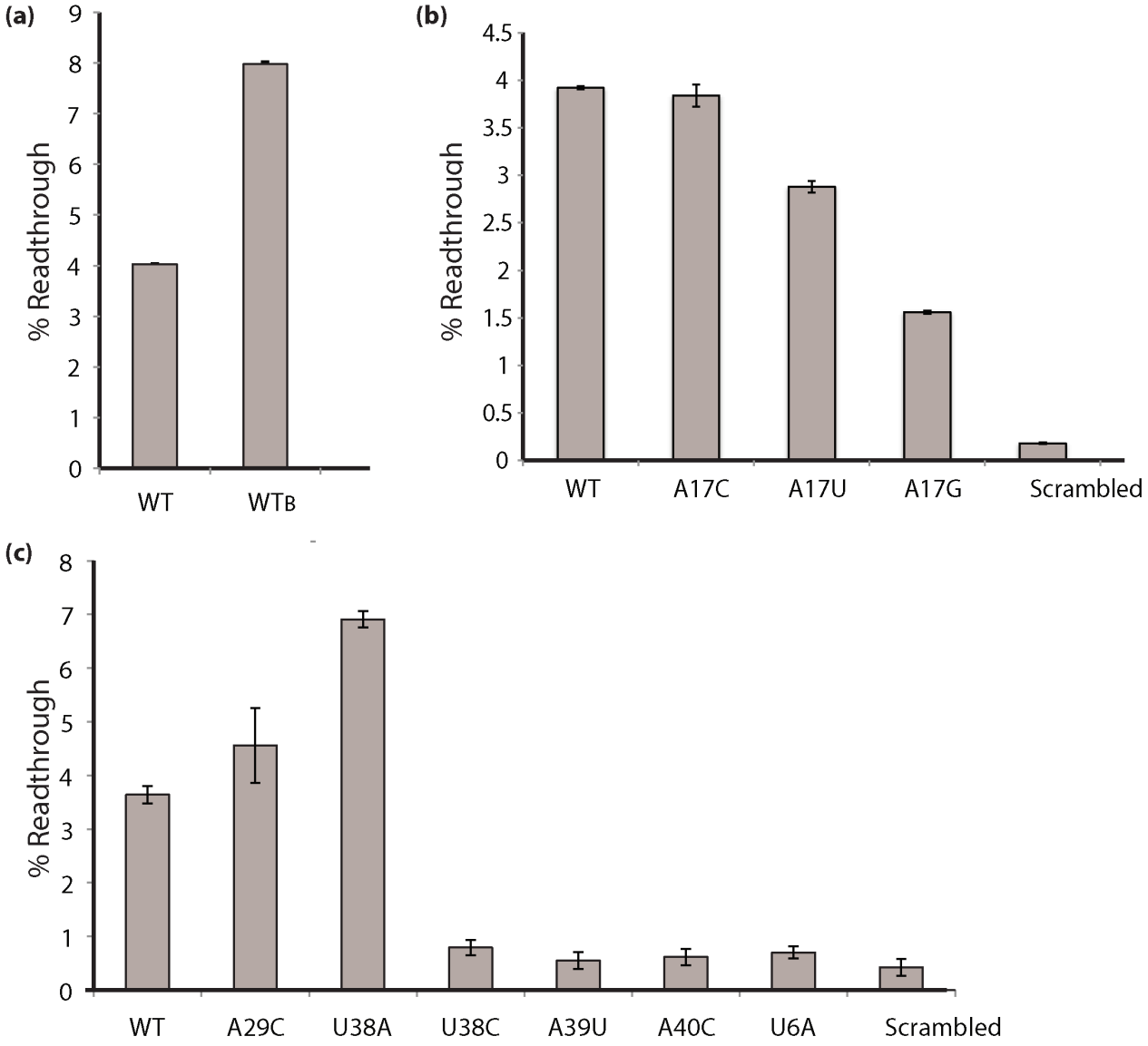


Supplementary Figure 4: Confirmation of MLV-PK assignments using specifically deuterated samples. Regions of 2D NOESY spectra for GU^H-MLV-PK **a,** and AC^H-MLV-PK **b,** samples showing sequential H1' to H8/H6 assignments for loop L2 (blue lines), stem S1 (solid lines) and stem S2 (dotted lines) residues at pH 7.5 (see Methods for assignment strategy). Coaxial stacking across the junction between residues G16/G52 and C24/C25 are indicated in bold red dotted lines in panels a and b, respectively. Although the tertiary interactions between stem S1 and loop L2 are not formed at high pH, typical adenine stacking interactions in the loop (blue line) are observed for residues A46-A48. **c,** Strip plots of NOE data showing the assignment of A17 by correlations to residues G16 and C51. The assignment is confirmed by residue specific labeling and mutational analysis (A17C and A17U).

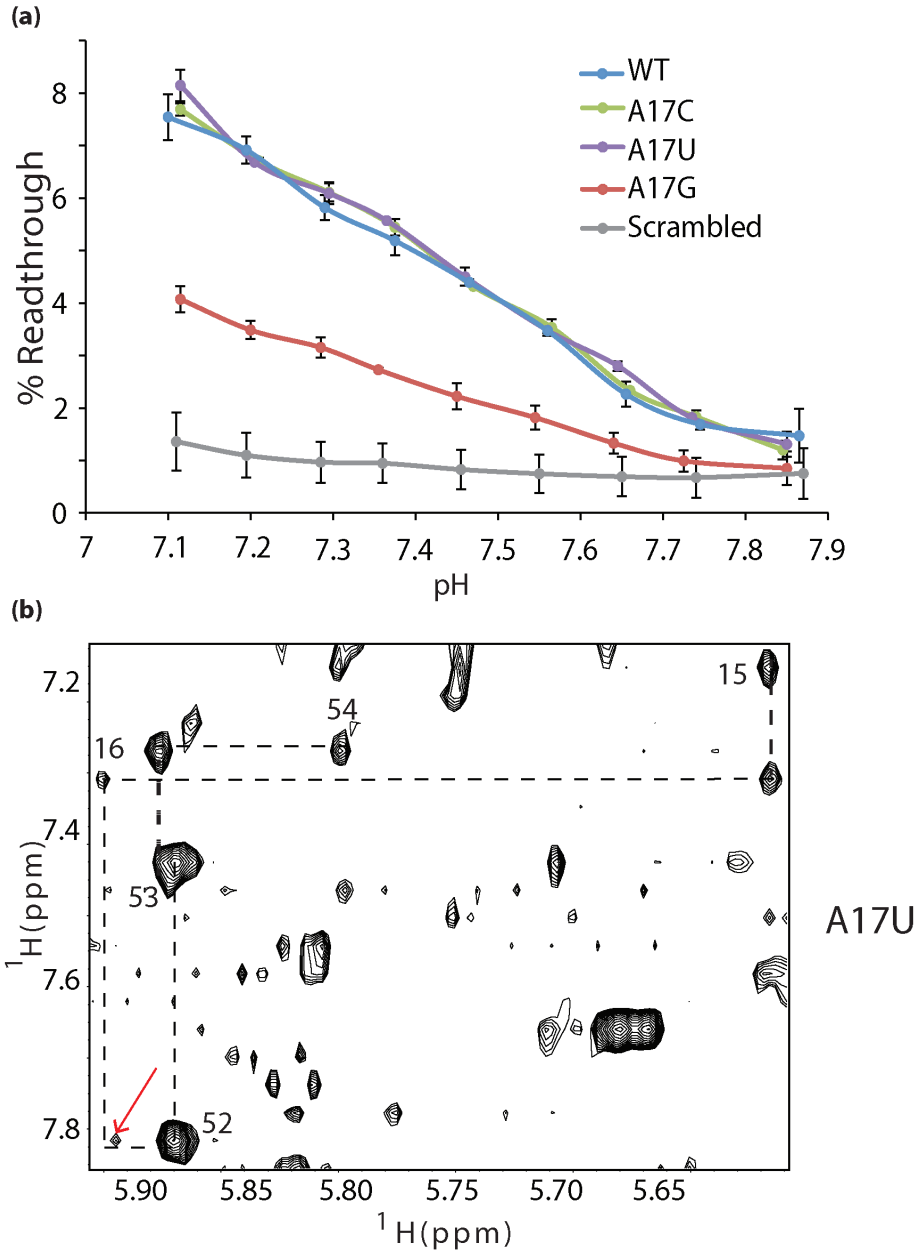


Supplementary Figure 5: Protonation-dependant formation of tertiary structure in MLV-PK. Adenosine C2 chemical shift changes in residue A14 **a,** and A17 **b,** during pH titrations. The decreasing intensity of the A14 C2-H2 resonance in $PK_{inactive}$ as the pH is lowered was used to determine the fraction of the inactive conformer. The intensities were normalized to the unperturbed A59 C2-H2 resonance between different datasets. The data were fit using a simple three-variable sigmoidal equation: f=a/(1+exp(-(x-x0)/b)) where x = pH, a = intensity at pH 8.5, b= intensity at pH 5.0 and x0 is the derived p*fold.*

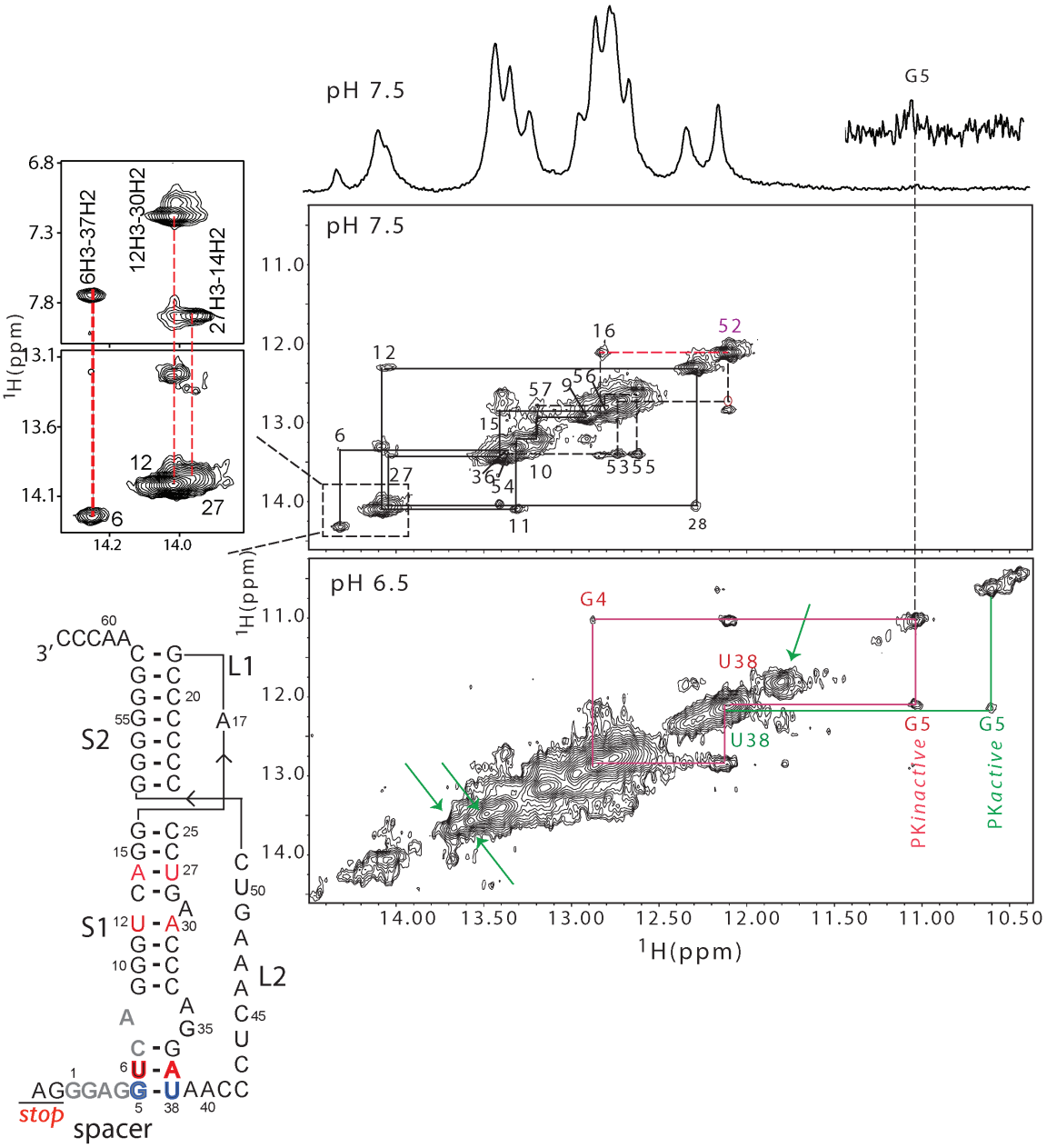
[Note: Titration at pH 5.0 results in chemical shift changes for all of the adenine C2-H2 groups, while the ribose and the aromatic C8-H8 groups stay the same, presumably due to the sensitivity of the pyrimidine ring to high proton concentration. However, since we are only using peak intensities for data fitting, the change in the chemical shift is not important for A14. While we use chemical shifts for A17, exclusion of this titration point during data fitting had no appreciable change in the derived p*Ka* value of A17.]



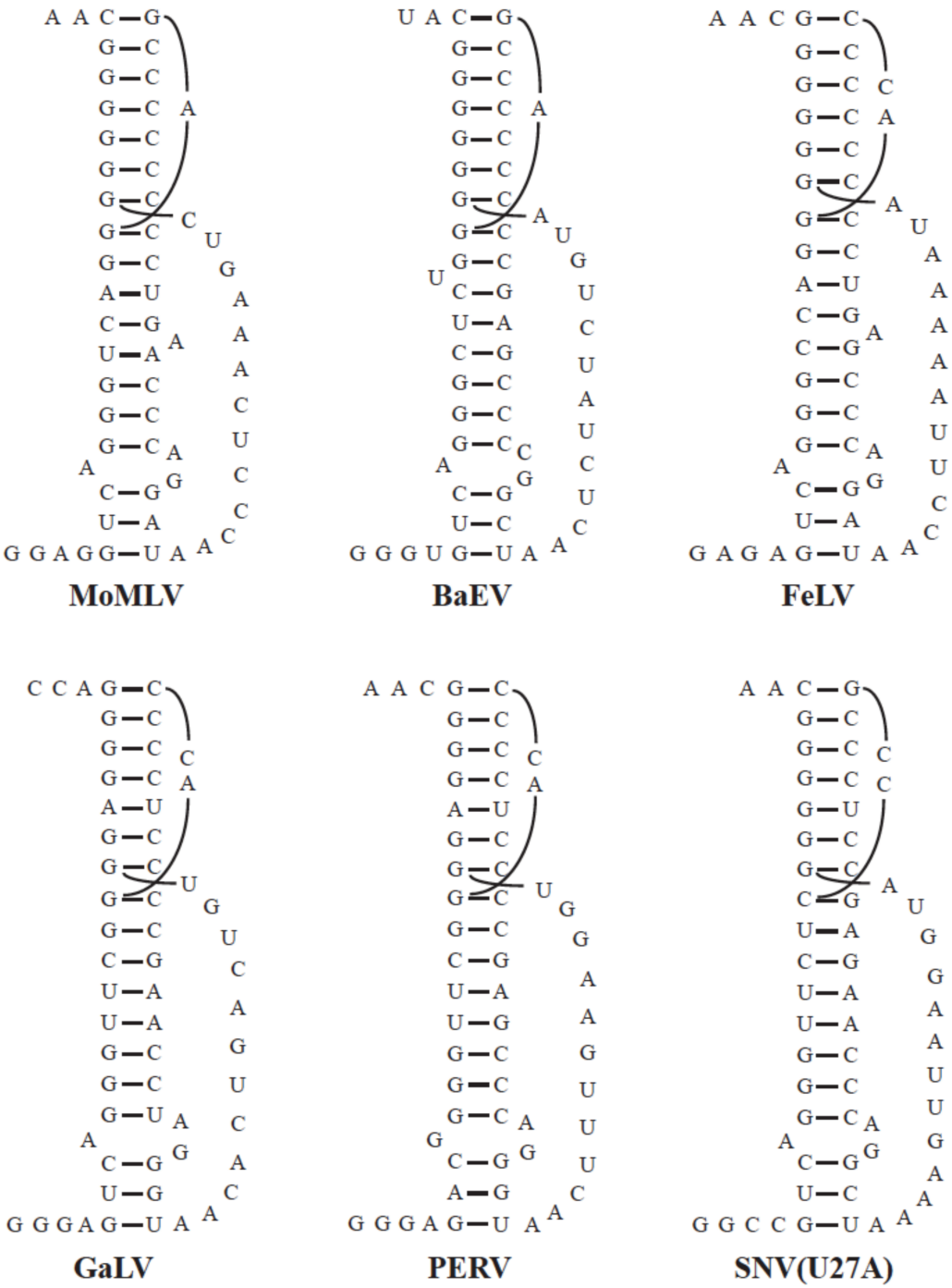
Supplementary Figure 6: *In vivo* functional assay for MLV-PK constructs. a, To test the *in vivo* levels of readthrough activity we used the previously described p2luc construct. We inserted the MLV-PK recoding sequence in two contexts (*SalI/BamHI,* WT and *BamHI/SacI,* WT_B*),* which result in 4% and 8% readthrough levels, respectively, suggesting that the flanking nucleotide sequences influence the observed readthrough values. While we do not understand the cause of this variation, the relative readthrough levels of mutants tested in both contexts are maintained. The *in vitro* and *in vivo* translation assays were thus carried out in the WT context unless otherwise indicated. **b,** *In vivo* readthrough assay results for A17 loop L1 mutants. **c**, *In vivo* readthrough assay results for S1-L2 turn mutants. For U38A we calculate a PK_{active}:$PK_{inactive}$ population distribution of 9:91 from the increased p*Ka* of 6.40, which would correspond to a ~50% increase in readthrough levels. This correlates with the ~80% increase observed *in vivo*.



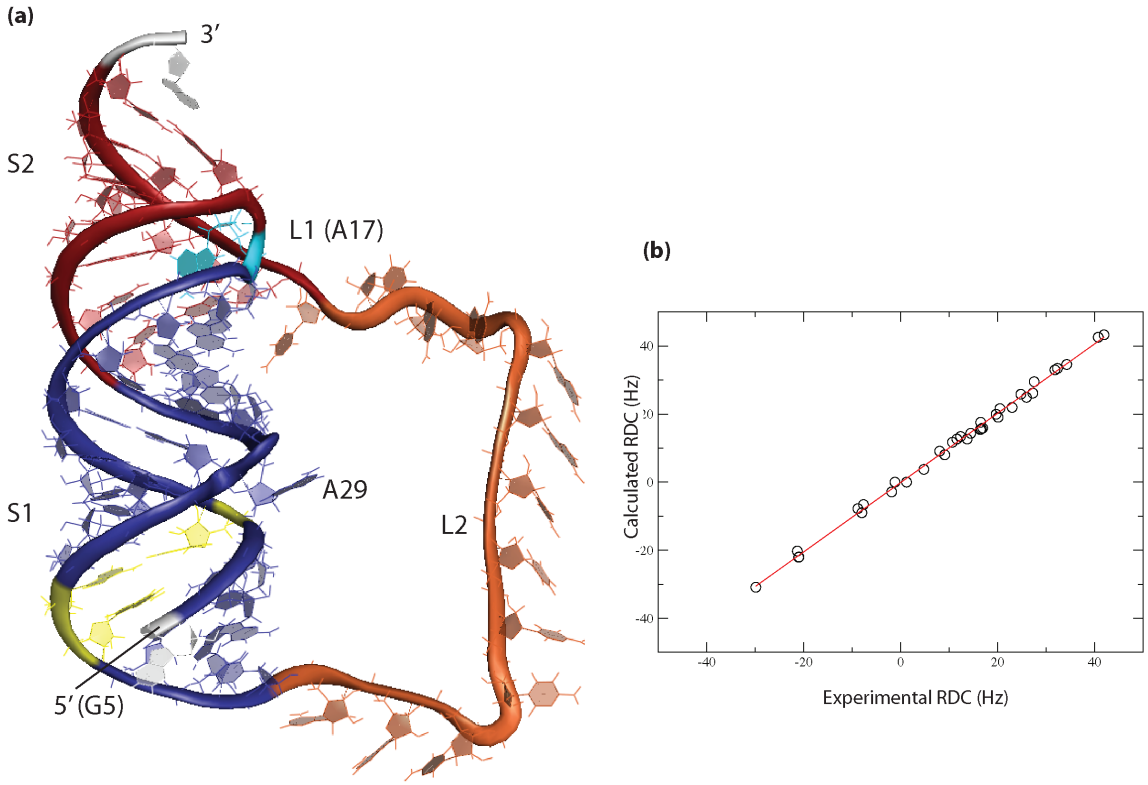
Supplementary Figure 7: MLV-PK contains multiple protonation sites. a, The effects of the translation pH on readthrough activity of MLV-PK A17 mutants *in vitro*. **b,** 2D NOESY aromatic to anomeric proton connectivities at pH 7.5 showing the loss of standard helical stacking interactions (red arrow) between G16 and G52 across the helical junction in A17U *even at high pH* indicating a preformed interhelical bend (see Supplementary Figs. 4 and 11). This is similar to changes that occur in the wild-type MLV-PK only upon protonation of A17 and thus explain the wild-type like activity for this mutant (see below). The G53 chemical shift is, however, unperturbed in the A17U mutant, which proves that the helical bend occurs without a triple base formation with C23-G53 base pair (see Supplementary Fig. 11). ***Note:*** Also see Supplementary Text.



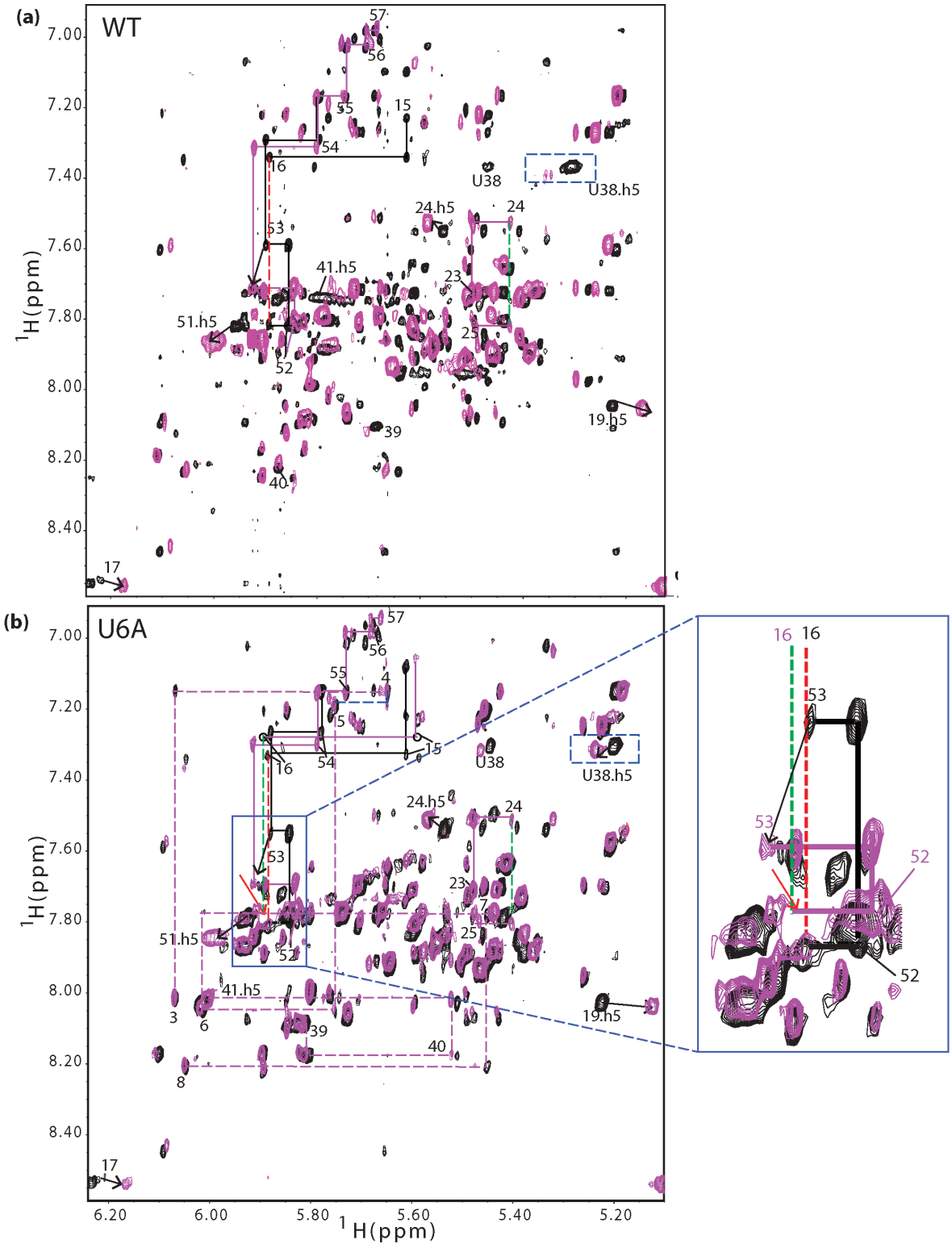
Supplementary Figure 8: NMR data confirm the extended stem S1 in MLV-PK. Complete imino proton NMR assignments were obtained for MLV-PK at pH 7.5. Black solid and dotted lines show connectivities for stems S1 and S2, respectively. GU wobble base pairs have typical upfield chemical shifts that require low pH conditions for observation in the NMR data. At pH 6.5, we observed a GU wobble in two conformations, one of which is also observed weakly at pH 7.5 (1D trace above with magnified inset) allowing for unambiguous assignment of GU wobble in PK_{active} (green) and $PK_{inactive}$ (red). In addition, the green arrows show imino resonances observed only in PK_{active}, again indicating a change in global conformation. Since AU Watson-Crick base pairs are readily assigned by intense NOEs from uridine H3 imino protons to adenine H2 protons (zoomed inset), our data confirm the extended stem S1 helix by assignment of U6 H3 NOE to A37 H2 and the imino-to-imino walk NOE to G36 H3.



Supplementary Figure 9: Secondary structure modeling of pseudoknots from the gammaretrovirus family. The extension of stem S1 with the 1 × 2 bulge can be modeled in other gammaretroviral readthrough signals and suggest that these motifs may be structurally conserved. All pseudoknots can be modeled with either a one- or two-base loop L1, each of which contains a nucleotide that can be protonated (A or C). In addition, each model contains a 1 × 2 internal bulge. The second stem S1 base pair (analogous to MLV U6-A37) shows variation from A-U, G•U, C•U and G•A. Interestingly, the terminal G•U base pair and the S1-L2 turn can be preserved by introducing the internal 1 × 2 bulge.

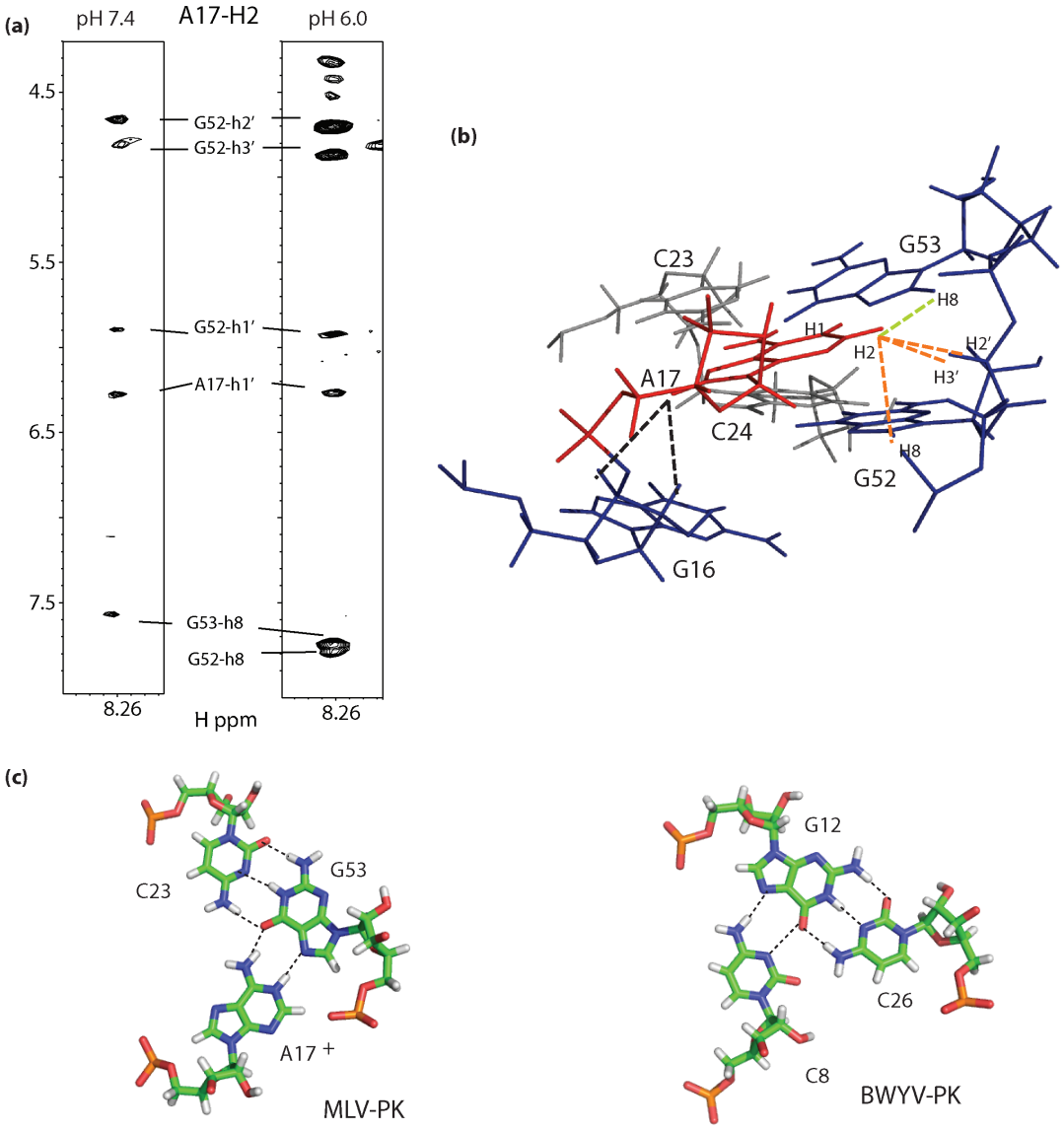


Supplementary Figure 10: Validation of PK$_{inactive}$ structure using residual dipolar coupling data. a, Representative AMBER refined structure of PK$_{inactive}$. For generating AMBER[30] structures, the 10 best CYANA structures were used. The refinement had 50,000 steps, with temperature increasing from 0 K to 500 K over the first 12500 steps, and remaining at 500 K over the next 32500 steps and then decreasing to 0 K over the next 5,000 steps. These calculations incorporated all upper limit restraints used in CYANA. The individual structures generated were then used for tensor fitting, and the above structure calculation process was repeated with the RDC restraints along with a final minimization that included 8000 steps. RDCs were collected as described previously[31]. **b,** Plot of observed *vs* calculated RDCs (R=0.998, slope=1.019) for the representative lowest energy structure.

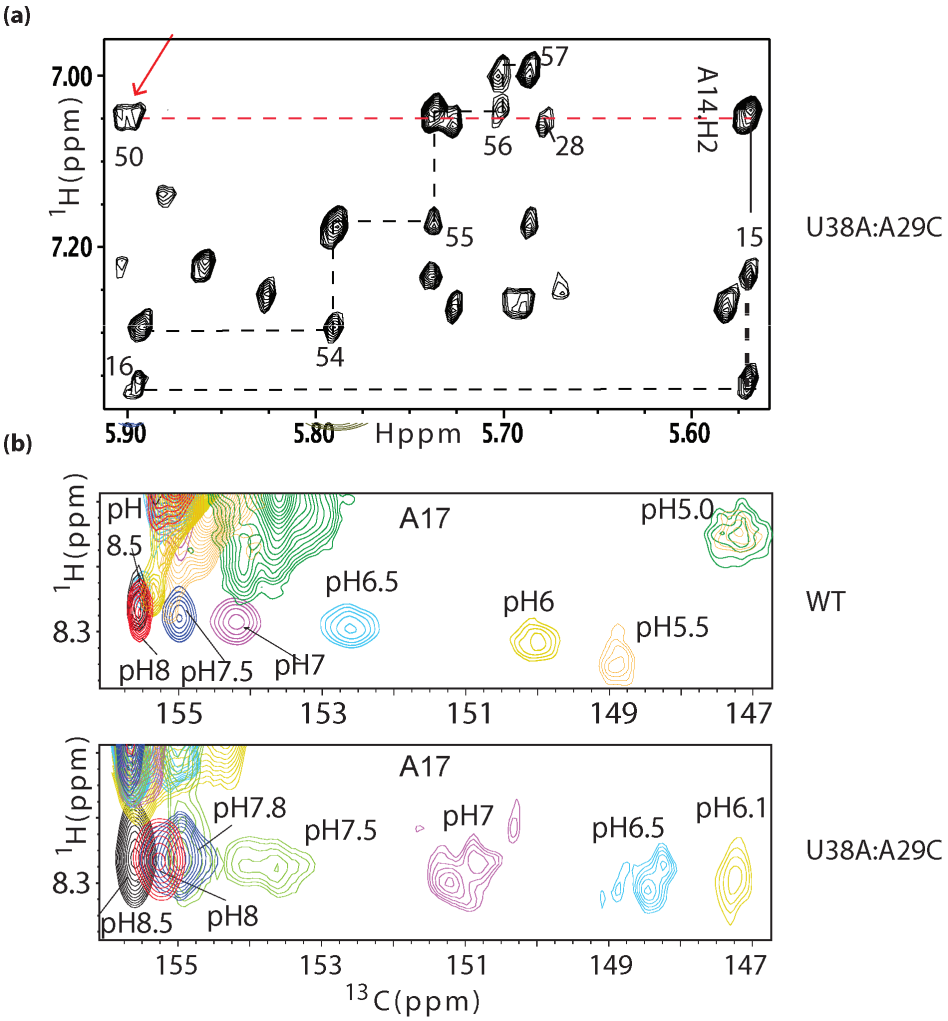


Supplementary Figure 11: Structural changes in wild-type and U6A MLV-PK constructs upon tertiary structure formation. Overlay of portions of 2D NOESY spectra for wild-type **a,** and U6A **b,** MLV-PK samples at pH 7.5 (black) and pH 6.5 (magenta), respectively. Chemical shift changes upon protonation of wild-type MLV-PK are seen throughout the structure. Line broadening due to chemical exchange is evident even at this intermediate pH (for example, residues 39-41), however, well-dispersed

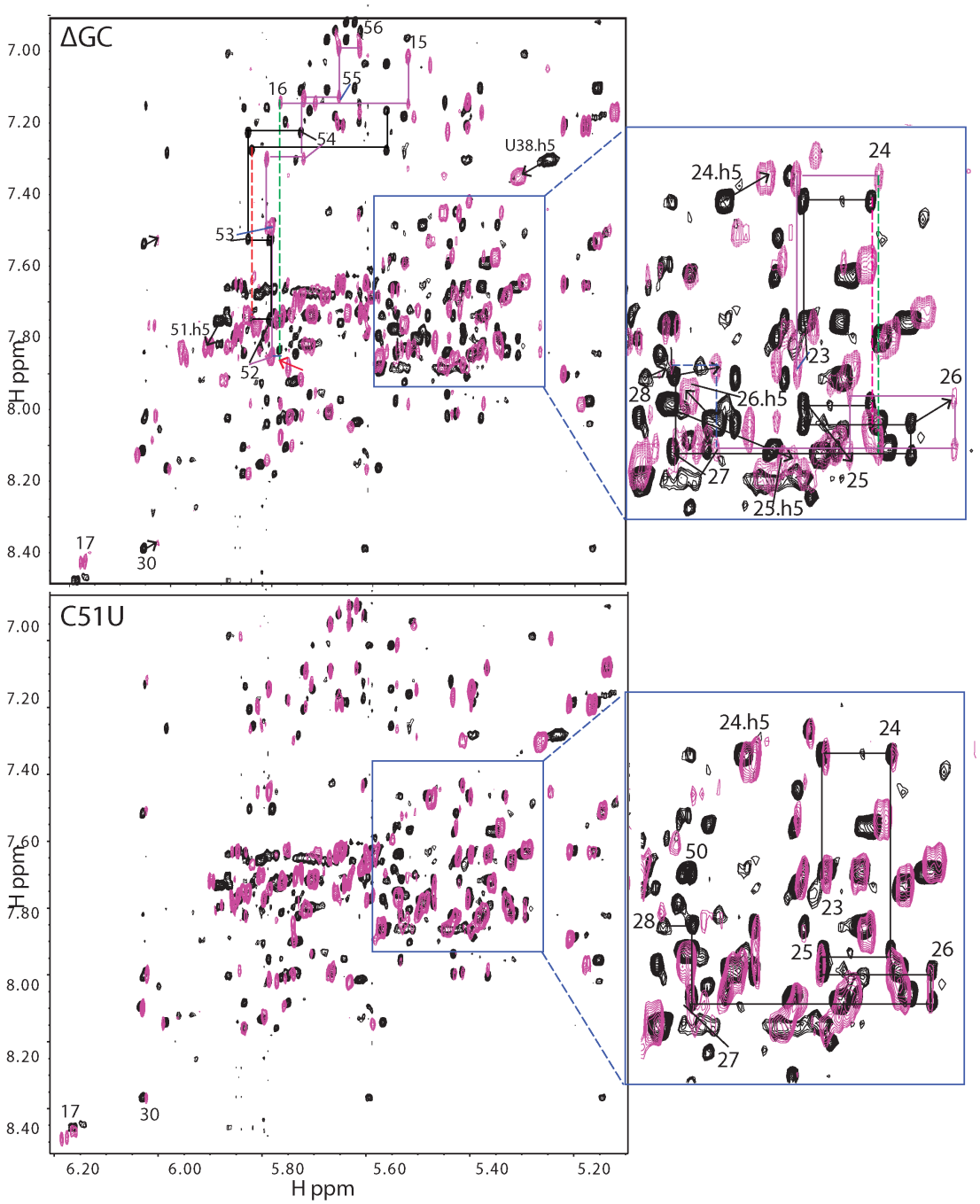
signals for residues in stem S2, and loop L1 (A17) allow for their unambiguous assignments in PK_{active} as indicated (magenta lines). Interhelical stacking (see Supplementary Fig. 4) is denoted by red dotted lines at high pH and the loss of stacking due to protonation is denoted by green dotted lines at low pH. The junction residues experience significant chemical shift changes that are denoted by black arrows. Residue G53 in stem S2 is particularly perturbed and experiences a 0.15 ppm downfield shift upon A17 protonation, indicating its role in triple base formation. Complete broadening of the S1-L2 turn (U38, black box) indicates the hypersensitivity of this resonance to change in tertiary structure. **b,** The inactive S1-L2 turn U6A mutant is capable of protonation at A17 as evidenced by similar changes in stem S2 and loop L1 as in wild-type MLV-PK, but, in contrast, cannot form the tertiary structure as indicated by lack of chemical shift perturbation and exchange broadening for stem S1 and loop L2. Consequently, complete assignment of the U6A mutant allowed for confirmation of changes that occur in stem S2 and at the junction and hence support the PK_{active} model. Interhelical stacking (see Supplementary Fig. 4) is denoted by red dotted lines at high pH and the loss of stacking due to protonation is denoted by green dotted lines at low pH (also see inset for a zoomed version). Interestingly, the U6A mutation maintains the extended stem S1 helical confirmation (dotted magenta lines), and the most dramatic changes are seen for residues A39-C41 in the S1-L2 turn. This suggests that an adenine at position 6 leads to an improper positioning of S1-L2 turn, which in turn inhibits tertiary structure formation.



Supplementary Figure 12: Spectroscopic evidence for the A^{+}17-C23:G53 base triple. a, Selected regions of 2D NOESY data for the A17 H2 and pH 7.4 (left) and pH 6.0 (right). These NOEs were used for structure calculations in CYANA and AMBER. Our structures consistently orient the base of A17 with the N1 facing N7 of G53 while the N3 is not positioned close to any proton acceptor. Modifying the CYANA and AMBER residue library to include a proton on the N1 position of the A17 was in agreement with the NOE data. **b,** NOE network mapped onto the PK$_{active}$ model supporting the formation of A17^{+}:G53-C23 base triple. In PK$_{active}$ protonated A17^{+} H2 has strong NOEs to G52 H8/H2'/H3' and intermediate NOEs to G53 H8/H2′/H3′. These NOE data are consistent with N1 protonation and not N3. In addition, the H8 resonance of the non-terminal G53 residue has a large chemical shift perturbation (see Supplementary Fig. 11), which may indicate hydrogen bond formation involving the adjacent N7 position. In A17U, which does not form a triple base the perturbation is not observed (see Supplementary Fig. 7). **c,** Hence, the triple base in MLV-PK, A17^{+}:G53-C23 (left) would be dependant on protonation at N1 and possibly include the A17^{+}:G53 amino-N6:caronyl-O6. Including the amino-N6:carbonyl-O6 hydrogen bond during structure calculations also agreed with the NOE data. In support of this model, BWYV (right) has a similar C^{+}G-C N3(+)-N7; amino-carbonyl, Watson-Crick triple base[18], and our data show that it is also an equilibrium dependant recoder (see Fig. 3c).



Supplementary Figure 13: NMR data for the MLV-PK mutants. a, Selected region of a 2D NOESY spectrum at pH 7.4 for U38A:A29C mutant showing the appearance of stem S1 (A14-H2)- loop L2 (U50.H1') NOE. In this mutant this region is predominantly in the folded conformation. **b,** Consequently, this mutant affords an elevated p*Ka* of 7:09 as evidenced by the A17 C2 chemical shift changes (see Fig. 3c for linear fit).



Supplementary Figure 14: Register shift of loop L2 in the ΔGC mutant. a, Overlay of portions of 2D NOESY spectra for wild-type (black) and ΔGC mutant (magenta) MLV-PK samples at pH 7.4. Partial assignments for the wild-type is shown in black lines and the interhelical stacking NOE is shown in green dotted lines. The ΔGC mutant was designed to engineer a bend at the junction. Assignments (magenta lines) confirm the absence of a GC base pair in stem S2. Most importantly, a majority of the resonances in stem S1 show a change in chemical shift (black arrows in inset) compared to the wild-type and indicate a register shift between stem S1 and loop L2 as a result of the engineered interhelical bend, see Fig. 4. **b,** For comparison, the junction C51U mutant (magenta) has a similar stem S1 and loop L2 orientation as the wild-type (black).